2016

Characterizing microRNA regulators of lung disease

BOSTON UNIVERSITY

SCHOOL OF MEDICINE

Dissertation

CHARACTERIZING MICRORNA REGULATORS OF LUNG DISEASE

by

CARLY BETH GARRISON

B.A., Boston University, 2009

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2016

Approved by

First Reader ____

Avrum Spira, M.D., M.Sc.
Professor of Medicine

Second Reader ____

Laertis Ikonomou, Ph.D.
Assistant Professor of Medicine

In the place I go there are things that I see

That I never could spell if I stopped with the Z.

I'm telling you this 'cause you're one of my friends.

My alphabet starts where your alphabet ends!...

So, on beyond Z!

It's high time you were shown

That you really don't know

All there is to be known.

-Dr. Seuss, On Beyond Zebra! (1955)

DEDICATION

For my parents,

You inspire me to reach for the impossible

And never accept “No” for an answer

Unless it’s regarding the 1967 Ford Mustang Shelby GT 500 I asked for…

ACKNOWLEDGMENTS

To my advisor – Avi: When I was close to leaving graduate school and all the endless frustrations associated with it, you took a chance and welcomed me into your lab. The first day we meet I remember you telling me that the most important part of any lab is the people. You have managed to build a section with only the best by your side and I'm honored to consider myself one of them. Thank you for your endless devotion to the science and to your students. You afforded me opportunities to learn and discover not only the science, but myself as well. I will forever be grateful for the opportunity I had to work, grow, and learn under your guidance.

To the rest of my committee – Marc, Laertis, Dave and Matt: The work in this dissertation would not have been possible if it wasn't for your guiding hands. You've afforded me the chance to learn and discover the science of microRNAs in ways I never would have thought of myself, ways that only come to light over cupcakes and committee meetings.

To my home away from home – The Spira/Lenburg Lab: Regardless of the question or problem, someone was always there to help. I have never worked with a more loyal and kind group of individuals. It has been an honor to call you all colleagues and learn from each and every one of you. Thank you for your endless patience and collaborative spirit.

To my Coauthors: The endless days and nights spent pouring over data, trying to understand what results mean, would have been maddening if not for your input. None of the work in this dissertation would exist if I didn't have a computational counterpart and the work we've done together is one thousand times better than anything I could have come up with on my own. Thank you for your insight.

To my friends, near and far: You've been there for every tear, every smile, every venting moment I've had on this journey and I can only thank you for allowing me the freedom to be myself.

To my family – Bruce, Raydene and Matthew: My endless passion for science is fueled by the support and encouragement you provide me. Throughout this journey you have been my sanity (which is a scary thought). The decision to remain in school for an advanced degree has been challenging at times, but all it takes is one visit to remind me that I'm not alone in any endeavor. I especially appreciate the care packages of cookies, penguins, and other surprising treats that reminded me that even though you're 3,000 miles away, you're with me every step of the way.

CHARACTERIZING MICRORNA

REGULATORS OF LUNG DISEASE

CARLY BETH GARRISON

Boston University School of Medicine, 2016

Major Professor: Avrum Spira, M.D., M.Sc., Professor of Medicine

ABSTRACT

Lung diseases are one of the leading causes of mortality and morbidity worldwide. Understanding these diseases at a molecular level remains a critical component to developing effective therapeutics. Previous work has shown that gene expression alterations play an important role in disease initiation, maintenance, and progression as well as serve as diagnostic tools in disease. However, much remains to be uncovered regarding the role that microRNAs play in both healthy and diseased lung tissue. This thesis seeks to utilize methods of bioinformatics, cell biology, and molecular biology to examine the effect of miR-4423 on lung epithelial cell differentiation (Aim 1), miR-424 on never smoker derived lung adenocarcinoma (Aim 2), and miR-34c isomiRs in interstitial lung disease (ILD) (Aim 3).

First, we examined the role of miR-4423 in lung mucociliary epithelium by employing the use of an air-liquid interface culture system, finding miR-4423 has an effect in ciliated cell differentiation and that a loss of miR-4423 is associated with cancer progression. These findings suggest that miR-4423's actions in airway epithelium differentiation may potentially provide a therapeutic role in lung cancer. Next, we validated transcriptomic differences between lung tumor tissues resected from never and

ever smokers. Specifically, miR-424, a predicted regulator of a large number of gene expression changes in never smoker lung adenocarcinoma, was found to regulate cell migration, potentially identifying a novel target and/or pathway for therapeutic action. Lastly, the function of microRNA isomiRs is relatively unknown. We validated miR-34c as upregulated in ILD and modulated both miR-34c and a miR-34c 5' isomiR in lung relevant cell lines to explore their differing biological roles. We found that they are capable of targeting differing mRNA, indicating an independent role for isomiRs in disease.

The studies contained in this dissertation offer valuable insight into the biology of microRNAs in the lung and how they might be employed as therapeutic targets for a number of common lung diseases. In addition, biological insights into the complexity of microRNAs in the lung highlight the need to better understand diseases influenced by microRNA expression and microRNA variants in regards to actionable therapeutics.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABC .... Avidin-Biotin Complex

AKT .... Protein Kinase B

ALI .... Air-Liquid Interface

ARAP2 .... ArfGAP with RhoGAP Domain, Ankyrin Repeat and PH Domain 2

ASCL1 .... Achaete-Scute Family BHLH Transcription Factor 1

ATCC .... American Type Culture Collection

BEGM .... Bronchial Epithelial Cell Growth Medium

BODE .... Body Mass Index, Airflow Obstruction, Dyspnea and Exercise Capacity

BSA .... Bovine Serum Albumin

CC10 .... Clara Cell 10 kDa Protein

cDNA .... Complementary Deoxyribonucleic Acid

CMV .... Cytomegalovirus

COPD .... Chronic Obstructive Pulmonary Disease

CRK .... V-Crk Avian Sarcoma Virus CT10 Oncogene Homolog

CRKL .... V-Crk Avian Sarcoma Virus CT10 Oncogene Homolog-Like

CT .... Cycle Threshold

DAB .... 3,3’ Diaminobenzidine

DAPI .... 4’,6-diamindino-2-phenylindole

DGCR8 .... DiGeorge Syndrome Critical Region Gene 8

dH_2O .... Distilled Water

DLCO .... Diffusing Capacity of the Lungs for Carbon Monoxide

DMEM:F12 .... Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12

DNA .... Deoxyribonucleic Acid

dsRNA .... Double Stranded Ribonucleic Acid

EGF .... Epidermal Growth Factor

eGFP .... Enhanced Green Fluorescent Protein

EGFR .... Epidermal Growth Factor Receptor

EMEM .... Eagle's Minimum Essential Medium

F-Actin .... Filamentous Actin

FACS .... Fluorescence-Activated Cell Sorting

FBS .... Fetal Bovine Serum

FC .... Fold Change

FDR .... False Discovery Rate

FEV1 .... Forced Expiratory Volume in 1 Second

FOXJ1 .... Forkhead Box J1

FOXP2 .... Forkhead Box P2

FVC .... Forced Vital Capacity

Gag-Pol .... Group-Specific Antigen-Polymerase

GAPDH .... Glyceraldehyde-3-Phosphate Dehydrogenase

GRAP .... GRB2-Related Adaptor Protein

GRB2 .... Growth Factor Receptor-Bound Protein 2

GSEA .... Gene Set Enrichment Analysis

H3Sp10 .... Specific Marker Ser-10-Phosphorylated Histone H3

HBEpC .... Human Bronchial Epithelial Cell

HEPES .... 4-(2-Hydroxyethyl)-Piperazine-1-Ethanesulfonic Acid

HTA .... Human Transcriptome Array

IDT .... Integrated DNA Technologies

IHC .... Immunohistochemistry

ILD .... Interstitial Lung Disease

IPF .... Idiopathic Pulmonary Fibrosis

IRB .... Institutional Review Board

K14 .... Keratin 14

KLK10 .... Kallikrein-Related Peptidase 10

KRAS .... Kirsten Rat Sarcoma Viral Oncogene Homolog

LCC .... Large Cell Carcinoma

LGRC .... Lung Genomics Research Consortium

LRP8 .... Low-Density Lipoprotein Receptor-Related Protein 8

LTRC .... Lung Tissue Research Consortium

MAPK .... Mitogen-Activated Protein Kinase

MDM2 .... Mouse Double Minute 2 Homolog

microRNA .... Micro Ribonucleic Acid

miRISC .... miRNA-Mediated Silencing Complex

miRNA .... Micro Ribonucleic Acid

MOI .... Multiplicity of Infection

MRE .... miRNA Response Elements

mRNA	Messenger Ribonucleic Acid
MSigDB	The Molecular Signatures Database
MUC5AC	Mucin 5AC
MUC5B	Mucin 5B
NHLBI	National Heart, Lung, and Blood Institute
NOTCH1	Notch 1
NOTCH4	Notch 4
NSCLC	Non-Small Cell Lung Cancer
NUSE	Normalized Unscaled Standard Error
PBS	Phosphate-Buffered Saline
PCA	Principal Component Analysis
PI	Propidium Iodide
PI3K	Phosphoinositide 3-Kinase
PIK3R3	Phosphoinositide-3-Kinase, Regulatory Subunit 3
Pre-miRNA	Precursor miRNA
Pri-miRNA	Primary miRNA
qPCR	Quantitative Polymerase Chain Reaction
qRT-PCR	Reverse Transcription Quantitative Polymerase Chain Reaction
RALA	Ras-Related Protein Ral-A
RHOA	Ras Homolog Gene Family, Member A
RHOC	Ras Homolog Gene Family, Member C
RLE	Run Length Encoding

RMA .... Robust Multi-Array Average

RNA .... Ribonucleic Acid

RNAse .... Ribonuclease

RNU44 .... Small Nucleolar RNA, C/D Box 44

RPM .... Reads Per Million

RPMI-1640 .... Roswell Park Memorial Institute Medium 1640

SCC .... Squamous Cell Carcinoma

SCLC .... Small Cell Lung Cancer

SE .... Standard Error

Tat .... Trans-Activator of Transcription

TGF-ß .... Transforming Growth Factor Beta

TKI .... Tyrosine-Kinase Inhibitor

TP53 .... Tumor Protein p53

UBC .... Ubiquitin C

UBE4B .... Ubiquitination Factor E4B

UTR .... Untranslated Region

VSV-G .... Vesicular Stomatitis Virus G

WDR63 .... WD Repeat Domain 63

CHAPTER ONE

Introduction

1.1 Airway Epithelium Differentiation

The airway epithelium is comprised of multiple cell types including basal, ciliated, goblet, secretory and neuroendocrine [1]. The varied cell types work together to protect the lower respiratory tract from inhaled toxins. Cells in the airway epithelium replenished by basal cell differentiation into the other cell types [2,3]. Cellular turnover is critical to ensure that damaged cells are removed and replaced, [4] a process important for preventing disease and ensuring the protective capabilities of the airway [5,6]. Disruption to the airway epithelial homeostasis leads to inflammation [7–11]. Chronic inflammation, caused from a wide variety of biological processes, can predispose an individual to lung disease such as cancer and interstitial lung disease[12–18]. An understanding of the mechanism of normal airway epithelial differentiation is a critical step in understanding events that disrupt this system, leading to lung disease.

The airway epithelium is constantly exposed to inhaled toxins including, carcinogens in tobacco smoke. Our laboratory and others have identified a "field of injury" throughout the respiratory tract and lung that is induced by cigarette smoke and/or lung disease and measurable by gene and microRNA expression profiling [19–21]. The "field of injury" is based on the concept that the entire airway is exposed to inhaled toxins, and while diseased tissue is in a distal part of the lung, the histologically normal airway cells react to an assault and can be used to measure distal disease-specific changes[22]. We have previously shown that tobacco smoke alters the gene expression of airway epithelial cells and that bronchial airway gene expression can serve as a minimally invasive biomarker for the detection of lung cancer [23–28]. Based on this body of evidence, the airway is not only important in protecting the lower respiratory tract but can also serve as an important diagnostic area for many smoking related lung diseases.

1.2 Lung Cancer

Lung cancer is the leading cause of cancer-related mortality globally with a 5-year survival rate around 15%[29]. Despite copious research in diagnostics and treatments[30,31], most patients are diagnosed at late stages, leading to poor prognosis[32]. Lung cancer is a heterogeneous disease with two main subtypes: small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC). While SCLC only comprises 15% of all lung cancer cases, it is a more aggressive form of lung cancer that originates in the upper airways, the bronchi, and grows quickly to metastasize to other parts of the body[33]. Most lung cancer cases are NSCLC (~85%) and can be broken down into further classification of adenocarcinoma (~50%), squamous cell carcinoma (SCC, ~30%) and large cell carcinoma (LCC, ~5%)[34]. SCC generally originates in the proximal airways and adenocarcinoma in the distal airways[35]. The wide range of locations of the subtypes of lung cancer within the lung lead credence to the concept that they arise from distinct cell types in the lung[36].

One of the most common causes of lung cancer is smoking tobacco cigarettes, a habit that is slowly on the decline in the United States[31,34,37]. Unfortunately, incidence of lung cancer in never smokers is on the rise with no strong implication for any specific environmental or genetic cause[38,39]. Studies show that there are clinical, genomic, and genetic differences in lung cancer cases presented in a never smoker compared to an ever smoker. Specifically, lung cancer in never smokers is predominantly adenocarcinoma compared to the percentages of other cancer types listed above that are found in ever smokers[38]. Hypothesized risk factors for never smoker lung cancer are varied, from environmental (for example diet, indoor air pollution, and cooking fumes) to genetic, compared to the prevalent risk factor of smoking in ever smokers[40]. Mutational spectra also vary depending on smoking status, with EGFR mutations more common in never smoker tumors and KRAS mutations in ever smoker tumors[40–42]. One of the more interesting, and as of yet unexplained features of never smoker lung cancer is that women are diagnosed with this variation of lung cancer at a 2:1 ratio to men[43]. There is a growing need to better understand the molecular biology of the numerous categories of lung cancer and leverage this knowledge to develop precision diagnostic and therapeutic options.

1.3 Interstitial Lung Disease

Interstitial lung disease (ILD) is the term used for a heterogeneous group of chronic respiratory disorders that are commonly associated with high morbidity and mortality with the most common ILD being idiopathic pulmonary fibrosis (IPF)[44,45]. The common feature of these diseases is a progressive remodeling of the alveolo-capillary barrier, essentially scarring of lung tissue. This scarring may cause progressive lung stiffness[45]. Clinically, these diseases present with shortness of breath and diffuse lung infiltration on CT scan[44].

The working concept of ILD pathogenesis shows repeated injury to alveolar epithelial cells and failure of the alveoli to correctly respond to the injury, leading to aberrant lung repair and progressive fibrosis. Our understanding of ILD is that there are both genetic and non-genetic factors contributing to alveolar incorrect repair[46,47]. As there are a number of diseases under the umbrella term of ILD, there are a number of potential causes and treatments. Most cases of ILD are diagnosed late in the disease progression and the majority of treatment options only slow progression of the disease because they treat the symptoms, not the underlying disease biology[44].

Many ILDs, such as IPF, have no effective therapies and gradually progress with worsening symptoms[48–50]. A better understanding of the molecular underpinnings of ILDs is necessary so that treatment options targeting the causal disease mechanism of fibrosis can be discovered.

1.4 MicroRNAs

In 1993 the first microRNA, lin-4, was discovered in *C. elegans*, but it was not until 2001 that the name microRNA was first used, after they were found to have regulatory roles[51,52]. MicroRNAs are between 20 and 25 nucleotides long and are endogenously expressed to regulate gene expression by binding to the 3’ UTR of a target gene. They can induce mRNA degradation or translational inhibition and have been found to be involved in nearly all biological processes[53].

MicroRNAs start in the nucleus with a single transcript that can be up to 1000 nucleotides long, called a primary microRNA (pri-miRNA). The pri-miRNA is then processed by the RNAse III enzyme Drosha and the RNA binding protein DGCR8, resulting in a pre-miRNA, a transcript 70-100 nucleotides long and in a hairpin structure[54]. The RanGTP-dependent dsRNA-binding protein Exportin-5 then transports the pre-miRNA to the cytoplasm where the enzyme Dicer processes it into two microRNAs, a 5p and 3p form[55]. One of these forms is commonly degraded leaving one microRNA transcript. In some cases both transcripts are active[56]. An Argonaute protein then sequesters the active microRNA into a microRNA-induced silencing complex (miRISC). The microRNA is then sent to its target mRNA to bind to the 3' UTR for translational degradation or repression[57]. Each microRNA has a 2-8 nucleotide "seed region" that is believed to be important for target recognition of a microRNA recognition/response element (MRE) in the 3' UTR of an mRNA. Some target genes may have more than one MRE, allowing regulation by more than one microRNA[58].

1.5 MicroRNAs in the Lung

Research over the past few years has shown microRNAs to play important roles in lung development and homeostasis. These studies began in mice and have slowly progressed to human airway tissue[59]. MicroRNAs play a conclusive role in many biological processes including apoptosis, proliferation and differentiation[12,53,60–65]. Disrupting normal microRNA function can lead to many lung diseases such as lung cancer [66,67].

The impact of microRNAs in lung cancer disease biology is studied more extensively than in other lung diseases, with interest in microRNAs potential usefulness for sub-classification of the different cancer types and as eventual therapeutic tools[68]. Many microRNAs such as miR-99b, miR-7515, miR-126, miR-21, miR-200c, miR-145, miR-107, miR-185, miR-101, miR-138 and miR-7 are found to be aberrantly expressed in lung cancer [69–72]. Studies of these microRNAs show them to be regulators of genes involved in pathways known to be connected to cancer such as miR-7 regulating PIK3R3 in the PI3K/AKT pathway [72]. Other lung diseases such as IPF are shown to have microRNA regulators such as let-7, miR-21, miR-155, and miR-29[73]. Fibrosis is modulated with the inhibition of miR-21 in a bleomycin model of lung fibrosis while perturbations of miR-29 and miR-155 are shown to have profibrotic effects[73]. While many microRNA have been studied in detail there are still numerous questions regarding the biology of the lung and aspects of lung disease biology that need to be explored with future microRNA studies.

1.6 MicroRNA Therapeutics

With the discovery of microRNAs came the flood of information regarding their role in gene regulation and therapeutic potential[74,75]. More than 60% of the human protein-coding genes have at least one conserved MRE along with a number of non-conserved binding sites, allowing for the control of the majority of protein-coding genes by microRNAs[76]. Although the potential of microRNA therapeutics is appealing, there are some important challenges to overcome before they can be efficient in a clinical setting. MicroRNAs are hydrophobic in nature and are predominantly negatively charged, making it difficult to deliver them across biological membranes[77]. MicroRNA inhibitors may not be able to distinguish between microRNAs within the same family, causing off-target effects[78]. Chemical modifications to make both inhibitors and mimics more stable also appear to induce toxicity[79]. Currently, microRNAs and microRNA inhibitors are delivered with viral and non-viral vectors, with lower toxicity and better results seeming to come from non-viral vectors such as lipid-based delivery systems[80,81].

Despite the challenges associated with microRNA therapeutics, they are a growing field of research for a number of companies worldwide and are responsible for the creation of many therapeutic companies including: Mello Biotech, Regulus Therapeutics, miRagen Therapeutics, InteRNA Technologies, Mirna Therapeutics, and Groove Biopharma[82–87]. These companies are researching microRNA therapies, both microRNA antagonists and microRNA mimics, for a number of diseases and health products. There have been microRNA clinical trials in the US. Examples include MRX34, a microRNA mimic, from Mirna Therapeutics, Inc. and Miravirsen, a microRNA antagonist, from Santaris Pharma[88,89]. MRX34 is a double stranded mimic of the tumor suppressor miR-34 that is packaged in a liposomal nanoparticle formulation[90]. Phase I clinical trials for MRX34 are recruiting patients for a myriad of cancers including: primary liver cancer, small cell lung cancer, lymphoma, melanoma, acute lymphoblastic leukemia, chronic lymphocytic leukemia, multiple myeloma, and myelodysplastic syndromes[91]. Miravirsen is an example of a microRNA inhibitor, in the form of an antisense oligonucleotide, currently in Phase II trials that inhibits miR-122 in the liver of hepatitis C virus patients[92]. With microRNAs being important and specific regulators of the cell, their use in disease therapy is highly appealing as long as they can be delivered to the cells of interest in a minimally toxic manner.

1.7 Dissertation Aims

The following aims seek to broaden our understanding of microRNAs and their dysregulation with lung disease by leveraging cell biology and molecular biology techniques to characterize microRNAs identified in RNA sequencing data as being differentially expressed in lung disease. These microRNAs include miR-4423, a previously unreported microRNA detected in the airway epithelium that appears to lose expression with the progression of cancer, miR-424, a microRNA that is differentially upregulated specifically in never smoker lung adenocarcinoma tissue, and miR-34c, which is upregulated in ILD and has a 5' isomiR that has the capacity to target Ras signaling genes. Together, these studies highlight the complexity of microRNA biology in the lung, provide insight into the specificity of microRNA regulation of gene expression, as well as identify potential therapeutic options for lung cancer and ILD.

1.7.1 Aim 1: MicroRNA 4423 is Involved in Lung Epithelial Differentiation and its Expression is Lost with Lung Cancer Progression

Proliferation and differentiation of lung epithelial progenitor cells are tightly controlled processes important for lung repair after injury[93]. These processes are regulated by many factors, including the highly conserved miR-34/449 family of microRNAs, which has previously been shown to play a role in ciliogenesis in multiciliated cells and to function as a tumor suppressor[94,95]. The goal of this study is to see if miR-4423, a primate-specific microRNA highly expressed in the airway epithelium and an inhibitor of anchorage-independent growth[96], also plays a regulatory role during airway epithelial differentiation. We utilized an air-liquid interface (ALI) to characterize the mucociliary differentiation of bronchial epithelial cells. We show that overexpression of miR-4423 during differentiation of bronchial epithelial cells at an ALI results in an increase in the number of FoxJ1 and β_{IV}-tubulin expressing cells, suggesting that this microRNA plays a role in promoting ciliogenesis in the airway. In addition, we show that miR-4423 expression decreases with increasing cancer progression. This data suggests that miR-4423 may serve alongside highly conserved microRNAs to confer robustness to the process of ciliogenesis in the airway of primates as well as serve as a tumor suppressor lost with lung cancer progression. Further studies are needed to evaluate the potential of miR-4423 as a therapeutic in lung carcinogenesis.

1.7.2 Aim 2: MicroRNA 424 is a Potential OncomiR in Never Smoker Lung Adenocarcinoma that Regulates Cell Migration

Although lung cancer is strongly associated with tobacco smoking, approximately 25% of all lung cancer patients worldwide are never smokers[97]. Previous work performing differential analysis of matched pairs of adenocarcinoma tumor and adjacent-normal tissue yielded 59 protein-coding RNAs (mRNA), 61 large non-coding RNAs, and 15 microRNAs whose expression were uniquely altered in never smokers. Integrating the mRNA and microRNA data created a regulatory network for never smokers. We hypothesize that lung adenocarcinoma in never smokers compared to former and current smokers (ever smokers) have distinct microRNA regulators that may be leveraged to identify personalized lung cancer treatment options. By analyzing the integrative regulatory network, we identified miR-424 as a potential regulatory element in never smoker lung adenocarcinoma. MiR-424 was up-regulated in never smoker tumor, highly connected as a regulatory hub in the network, and has been previously described as an oncomiR in other cancers[98,99]. Differential expression of miR-424 was validated by a secondary platform, qRT-PCR, and found to have no association with EGFR mutational status. Interestingly, miR-424 is more highly expressed in the small airways of never smokers compared to ever smokers, a result requires additional studies to explore the potential of a airway biomarker for never smokers. Knockdown of miR-424 in never smoker adenocarcinoma cell line H2085 followed by gene expression array profiling revealed perturbations in many cancer hallmark pathways, including a number of migration related pathways. Strikingly, the functional knockdown of potential oncomiR miR-424 in the H2085 cell line significantly reduced migration by a scratch assay and the overexpression of miR-424 in the non-migratory cell line SK-LU-1 significantly increased migration. Further characterizing these discoveries will help elucidate the molecular underpinnings that differ between lung adenocarcinoma in smokers and in never smokers. These studies may ultimately lead to novel therapeutic targets and airway biomarkers for lung adenocarcinoma in never smokers.

1.7.3 Aim 3: MicroRNA 34c 5' IsomiR Regulates Ras Signaling Genes in Interstitial Lung Disease

The molecular mechanisms of the pathogenesis of ILD are still being uncovered[100]. As part of the Lung Genomics Research Consortium (LGRC), we previously performed small RNA sequencing on lung tissue from patients with and without ILD. Differential expression analysis showed miR-34c as highly induced in ILD tissue, with reads mapping to not only the canonical microRNA sequence, but that of 5' and 3’ isomiRs as well. The purpose of this study was to validate differential expression of miR-34c in ILD compared to control tissue as well as characterize miR-34c and its 5’ isomiR by exploring their direct targets that are shared by both forms as well as direct targets that are specific to each variant. Predicted targets of miR-34c and the miR-34c 5’ isomiR were enriched for genes involved in Ras signaling, a group of genes that are involved in pathways such as the EGFR pathway, whose dysregulation is described as conferring susceptibility to ILD[101,102]. By transiently transfecting lung fibroblast cells, we show that the 5’ isomiR of miR-34c targets GRB2, an important early member of the RAS/EGFR pathway. In addition, the canonical miR-34c microRNA fails to downregulate GRB2, indicating specificity for the 5’ isomiR. These data highlight the effect a small variation in the canonical microRNA has on the ability of a microRNA to regulate gene expression as well as introduces a miR-34c 5’ isomiR as a potential therapeutic target for targeting the RAS/EGFR pathway.

CHAPTER TWO

MicroRNA 4423 is Involved in Lung Epithelial Differentiation and its Expression is Lost with Lung Cancer Progression

2.1 Introduction

The airway epithelium provides a barrier against inhaled toxins, frequently undergoing cell turnover to replace those cells that are damaged with the rate of turnover depending on the rate of injury [5,13]. Repair occurs by multipotent basal cells migrating to the denuded area, proliferating and differentiating into the other airway cell types, such as ciliated, secretory and goblet cells [103]. Ciliated cells in the airway epithelium are terminally differentiated [104,105] and as such it is believed that growth arrest is necessary for ciliated cell differentiation.

MicroRNAs have been shown to play roles in many biological processes including apoptosis, proliferation and differentiation[12,53,60–65]. While microRNAs, such as miR-449 and miR-34, have previously been shown to play important roles in airway epithelial differentiation and development [94,106,107], the biological function of miR-4423, a recently discovered primate specific microRNA in the airway epithelium[96], is unclear. Previous work shows that both the 5p and 3p forms of miR-4423 are expressed in organs with a mucociliary epithelium (Figure 2.1A), which include the airway and fallopian epithelium[96]. In addition, miR-4423 is expressed throughout the respiratory tract, with positive *in-situ* staining in the airway of the trachea, mainstem bronchus, and second generation bronchus (Figure 2.1B)[96]. Building upon the data previously generated, we seek to evaluate the role of miR-4423 in differentiation of the normal airway epithelium using *in-vitro* manipulation of miR-4423 expression levels in an ALI system.

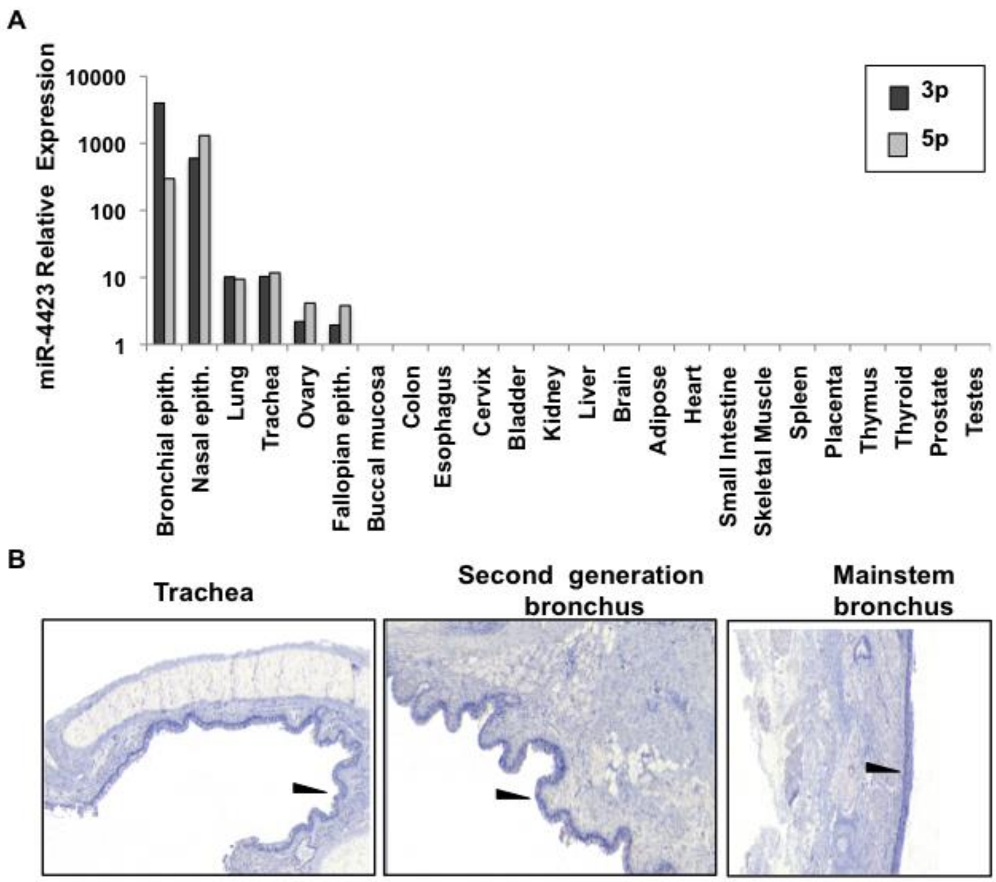

Figure 2.1: The expression of miR-4423 is primarily restricted to mucociliary epithelium. Adapted from Perdomo et al. (**A**) Expression of both forms of miR-4423 across 24 human tissues is detected in the respiratory tract (lung, trachea, and nasal and bronchial epithelia), ovary, and fallopian tube epithelium. (**B**) By in situ hybridization, miR-4423 is expressed in the epithelium of the trachea, mainstem bronchus, and second generation bronchus. Arrowheads point to the regions with positive staining.

Disruption to the process of airway repair, including disruption of microRNAs can lead to many lung diseases such as lung cancer [66,67]. In addition to being expressed in the airway epithelium, miR-4423 also inhibits tumor growth *in-vitro* and *in-vivo*[96]. Previous studies show that miR-4423 overexpression in "sensitive" cell lines prevents colony growth in soft agar (Figure 2.2A) and that in the H1703 SCC cell line, overexpression of miR-4423 results in a decrease in xenograft tumor volume as well as tumor weight (Figure 2.2B)[96]. Phosphorylated E-cadherin staining in the H1703 xenograft tumor tissue, shows positive staining in the miR-4423 overexpressing tumors, but not in the controls (Figure 2.2C), indicating an increase of cell-to-cell adhesions in the tumor and a possible mechanism for which miR-4423 is inhibiting tumor growth[96].

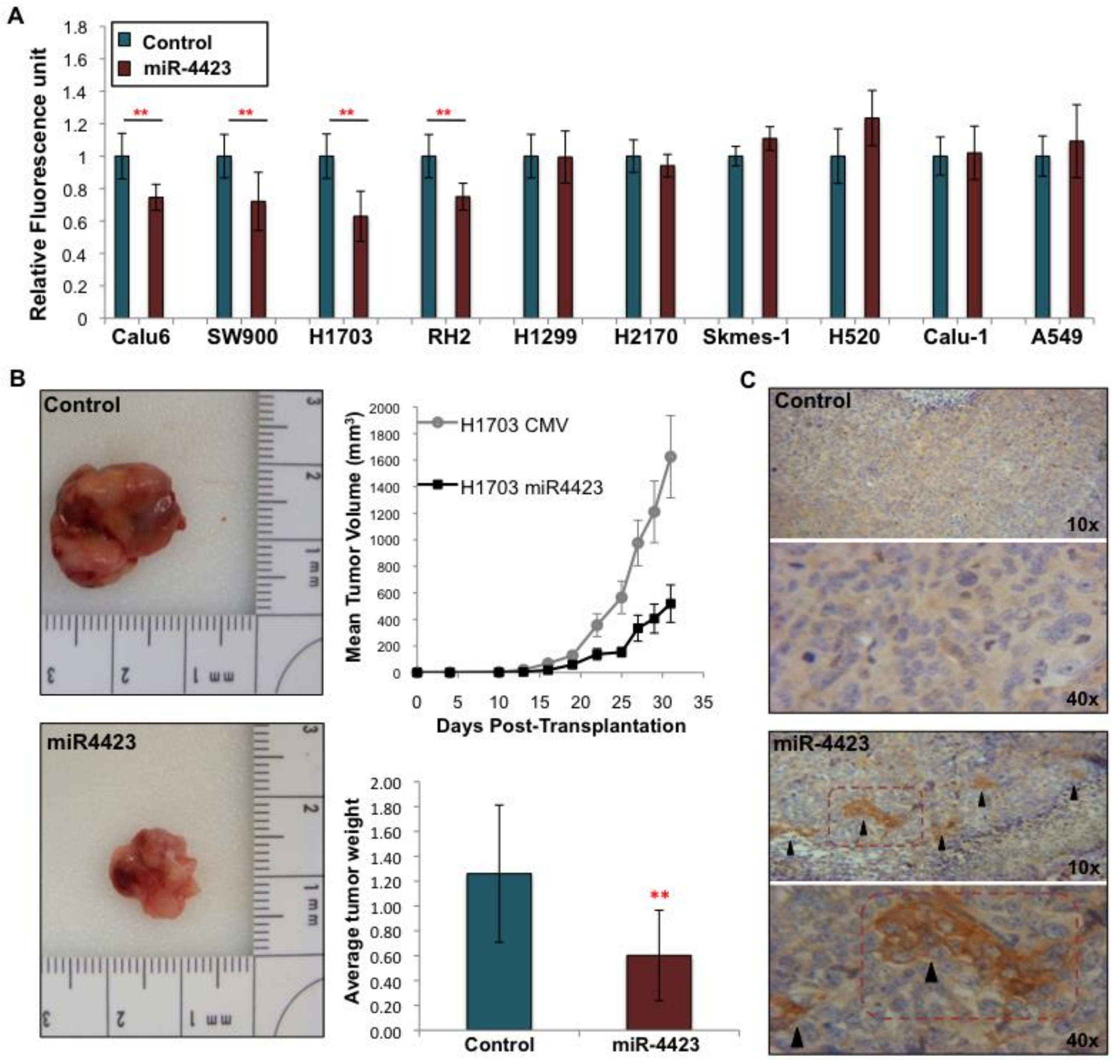

Figure 2.2: MiR-4423 inhibits lung cancer anchorage-independent growth in vitro and tumor growth in vivo. Adapted from Perdomo et al. (**A**) Soft agar assays were performed in the indicated cell lines stably transfected with either a vector that overexpresses the miR-4423 precursor or the empty parent vector as a negative control (n=10). Overexpression of miR-4423 in four of the cell lines tested decreases the number of colonies formed in soft agar (CalU6, $p = 2.8 \times 10^{-7}$; SW900, $p = 2.4 \times 10^{-5}$; H1703, $p = 7.4 \times 10^{-9}$; RH2, $p = 0.001$). Error bars indicate SE, and p values were determined using Student t test. (**B**) H1703 (SCC) cells stably overexpressing miR-4423 or a control (1×10^6) were injected subcutaneously into the backs of NSG mice (seven mice per group). Tumors derived from miR-4423 overexpressing cells were growth-suppressed relative to the control-derived tumors as shown in representative photographs of the tumors (Left), tumor volume over time ($p = 1.55 \times 10^{-11}$) (Upper Right) and tumor weight ($p = 0.01$) (Lower Right). (**C**) Phosphorylated E-cadherin staining was performed in miR-4423-overexpressing tumors and controls (H1703). MiR-4423-overexpressing tumors exhibited focal areas of positive membrane phospho-E-cadherin staining consistent with the presence of adherens junctions (Lower). Phospho-E-cadherin staining was not observed in control tumors (Upper). Arrows point to regions of positive staining.

Following the studies showing that miR-4423 is present in the airway epithelium and its expression in lung cancer cell lines inhibits tumor growth, we sought to explore the role of miR-4423 in the differentiating airway epithelium and its expression with the progression of lung cancer.

2.2 Results

2.2.1 Expression of miR-4423 in an ALI culture

To better understand the expression of miR-4423 in the airway epithelium, an ALI culture system was used and HBEpCs were grown through a differentiation protocol. Looking at miR-4423 expression over time, both forms of the microRNA appear to be expressed starting around day six of an ALI culture with increased expression as the culture continues to differentiate (Figure 2.3).

By measuring the expression of airway cell type markers *FOXJ1* (ciliated cells), *K14* (basal cells), *ASCL1* (neuroendocrine cells), *MUC5AC* (goblet cells), and *CC10* (secretory cells) we see that basal cells are continuously present in the culture with other cell type markers arising early in the differentiation process (Figure 2.4). Markers of ciliated cells appear to have the highest correlation with miR-4423 expression. *WDR63*, the transcript for a protein found in the inner dynein arm of cilia, is highly correlated with both miR-4423-5p (R=0.89) and miR-4423-3p (R=0.95). *FOXJ1*, which codes for a transcription factor in the nucleus of ciliated cells, is similarly correlated with miR-4423-5p (R=0.9) and miR-4423-3p (R=0.93; Figure 2.5). Both these genes and arms of miR-4423 are induced around day six and have increased expression throughout the differentiation process.

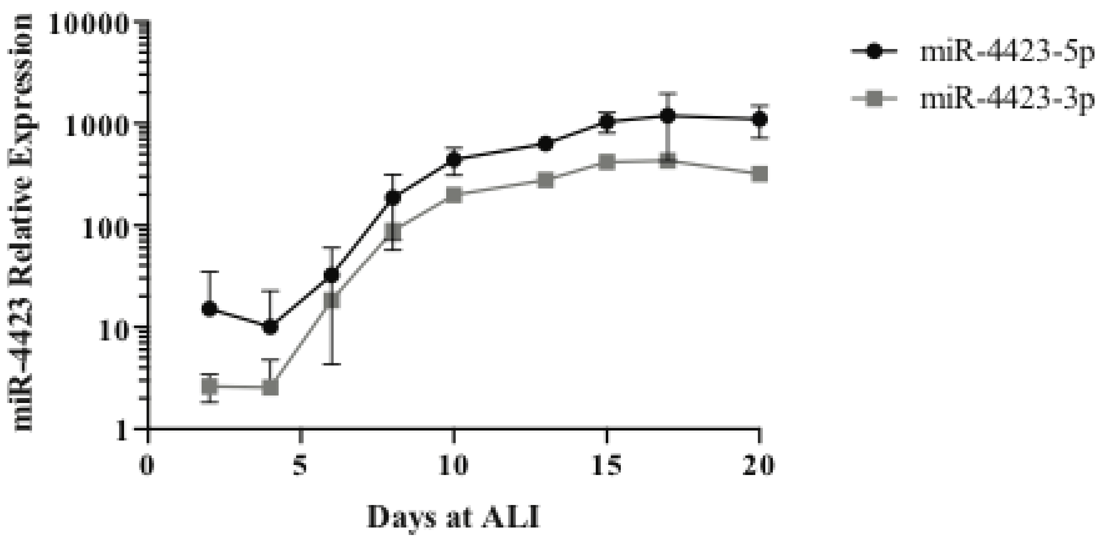

Figure 2.3: MiR-4423 expression during normal differentiation at an ALI. Basal cells underwent an ALI differentiation protocol over a period of 20 days with (n=2) inserts collected at each time point. Expression of miR-4423-5p and miR-4423-3p first appears around day six of differentiation and continues to increase during the protocol duration, reaching a plateau around day 15.

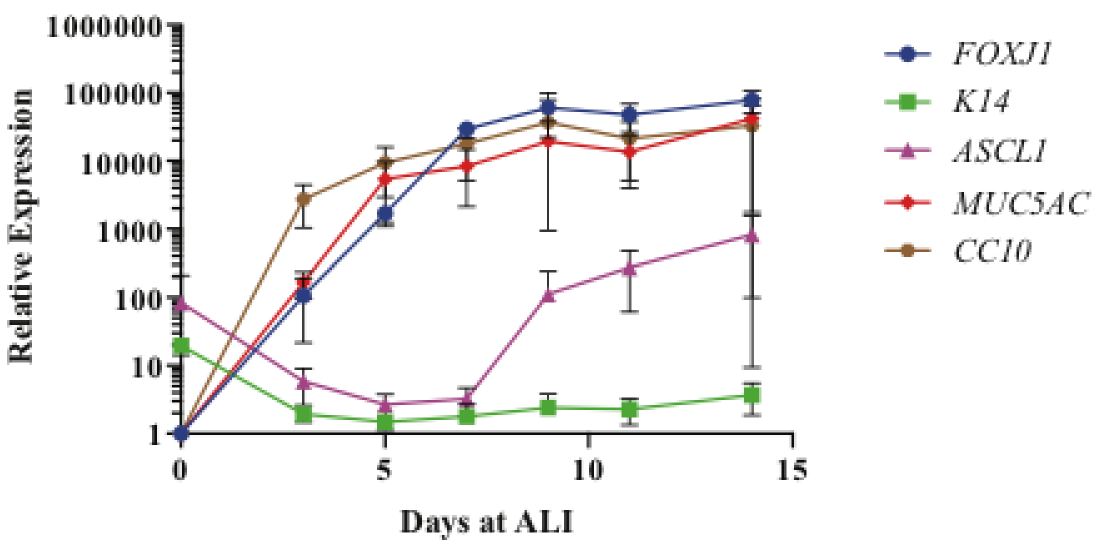

Figure 2.4: Cell type markers for all large airway cell types can be measured at an ALI. Basal cells underwent an ALI differentiation protocol over a period of 14 days with (n=3 to n=6) inserts collected at each time point in two experimental batches. Expression of *FOXJ1* (ciliated cells), *K14* (basal cells), *ASCL1* (neuroendocrine cells), *MUC5AC* (goblet cells), and *CC10* (secretory cells) is measurable during the differentiation process, highlighting the utility of an ALI as a biological model of the airway epithelium.

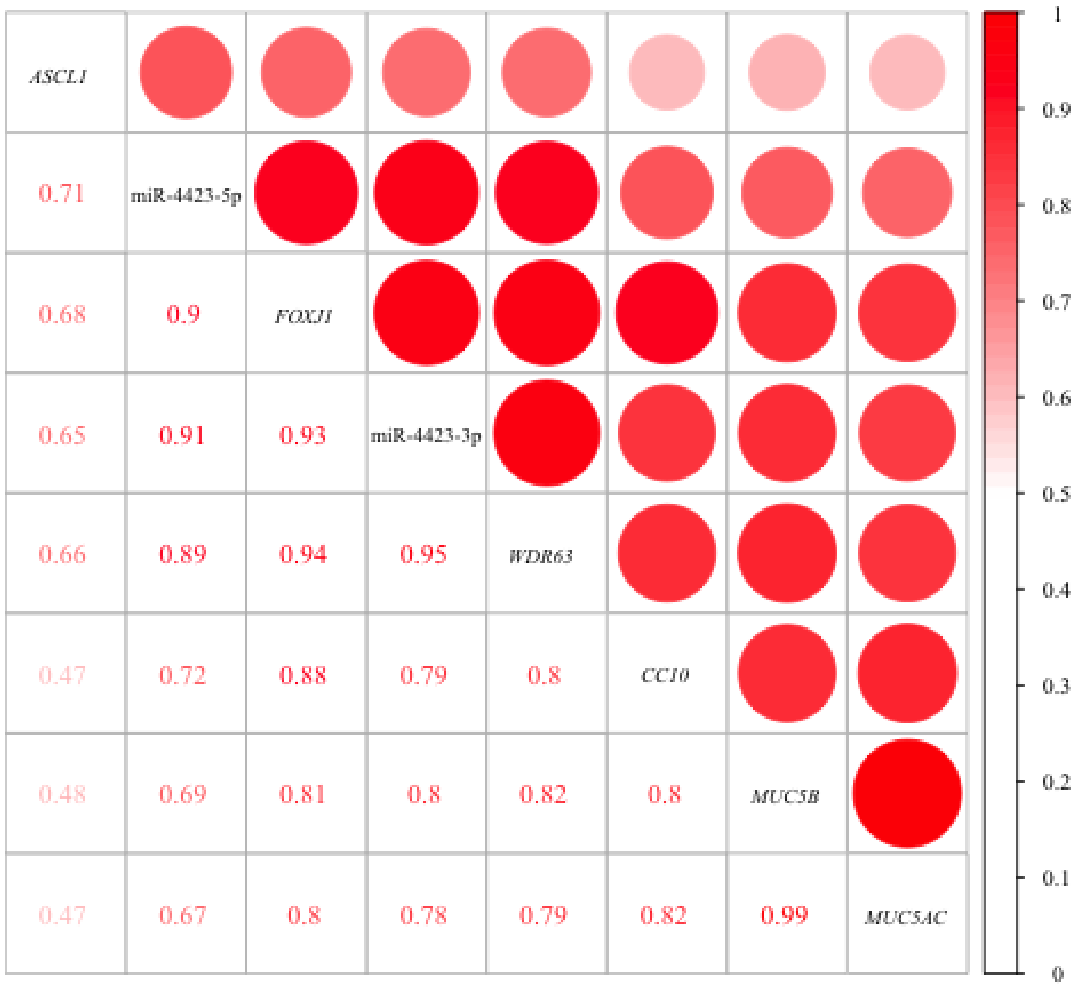

Figure 2.5: Cell type markers of ciliated cells most highly correlate with miR-4423 expression at an ALI. Pairwise spearman correlations between miR-4423-5p, miR-4423-3p, and the airway cell type markers *ASCL1* (neuroendocrine cells), *FOXJ1* & *WDR63* (ciliated cells), *CC10* (secretory cells) and *MUC5B* & *MUC5AC* (goblet cells) show the highest correlation between miR-4423 and the ciliated cell markers. Correlations are shown using a hierarchical clustering matrix with higher correlation coefficient indicated by the color red.

2.2.2 Constitutive miR-4423 overexpression induces ciliated cell production

HBEpCs infected with a constitutively active miR-4423 lentivirus show high levels of both miR-4423-5p (Figure 2.6A) and miR-4423-3p (Figure 2.6B) throughout all time points of the ALI differentiation process. All gene expression is normalized to eGFP expression to account for any transduction efficiency differences between batches. Measuring markers of the cell types in an airway epithelium after basal cells overexpressing miR-4423 differentiate at an ALI, we see that *FOXJ1* expression is increased with miR-4423 overexpression compared to control cultures ($p<0.005$; Figure 2.7). Looking closely at FOXJ1 in the ALI inserts from the miR-4423 overexpression experiments by staining for FOXJ1 with immunofluorescence, we see that there are an increased number of cells expressing the ciliated cell nuclear marker at day 11 of the differentiation process (Figure 2.8). Staining for β_{IV}-tubulin, marking the cilia of ciliated cells, we see an increased number of cells in miR-4423 overexpressing ALI cultures staining positive (Figure 2.9) at day 11. These results indicate that miR-4423 expression induces ciliated cell production during differentiation at an ALI.

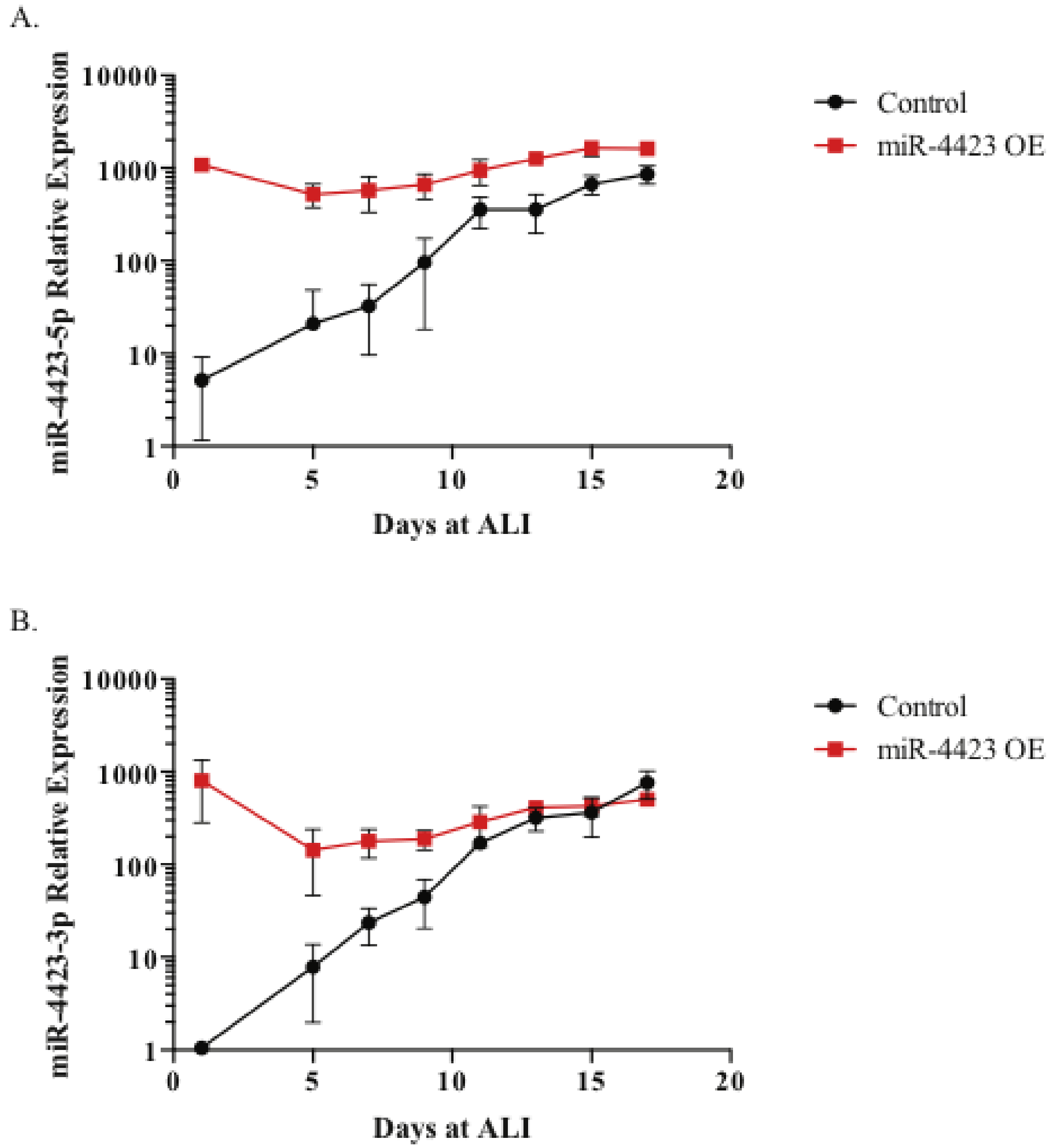

Figure 2.6: Overexpression of miR-4423 is measurable at an ALI. Basal cells infected with a lentivirus constitutively overexpressing miR-4423 successfully express (**A**) miR-4423-5p and (**B**) miR-4423-3p at a high level during the ALI differentiation protocol compared to control basal cells infected with an empty vector. Two ALI batches were normalized together with a total of n=2 to n=4 samples per time point.

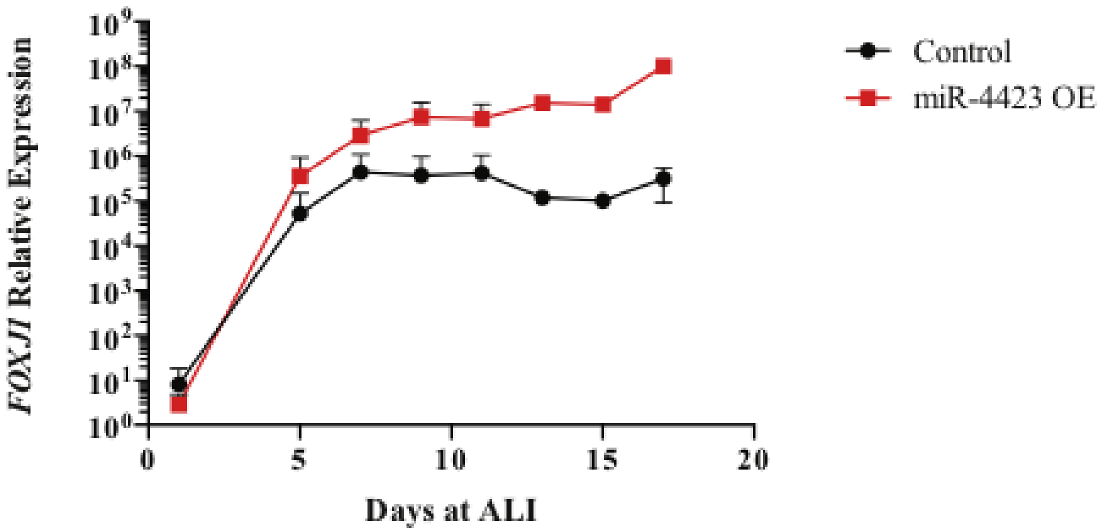

Figure 2.7: Ciliated cell marker *FOXJ1* expression is increased with overexpression of miR-4423. Basal cells infected with lentivirus constitutively overexpressing miR-4423 show a higher overall expression of *FOXJ1* during the differentiation protocol at an ALI ($p<0.005$). Two ALI batches were normalized together with a total of n=2 to n=4 samples per time point.

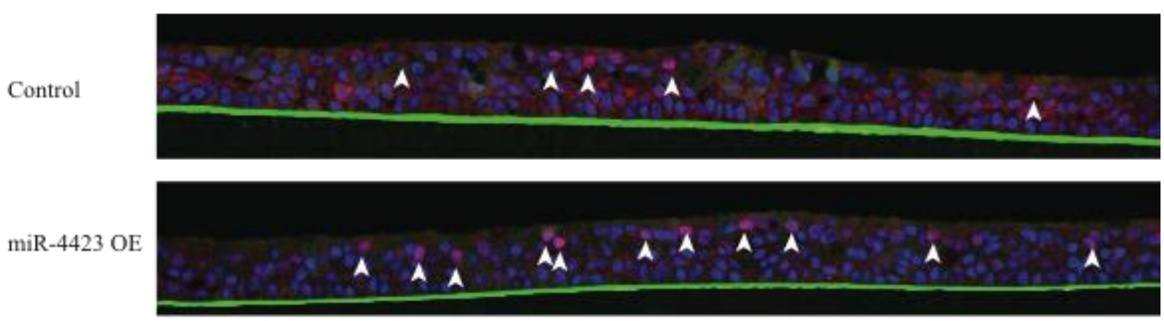

Figure 2.8: Ciliated cell marker FOXJ1 staining is increased with overexpression of miR-4423. Basal cells infected with lentivirus constitutively overexpressing miR-4423 show more cells with nuclear staining of FOXJ1 at day 11 of the differentiation protocol at an ALI compared to control. DAPI is shown in blue, FOXJ1 is shown in red, and green is present from autofluorescence of the ALI membrane as well as from the lentivirus used to infect the cells. Arrows point to regions of positive co-localization of DAPI and FOXJ1 staining. Images are at 10x magnification.

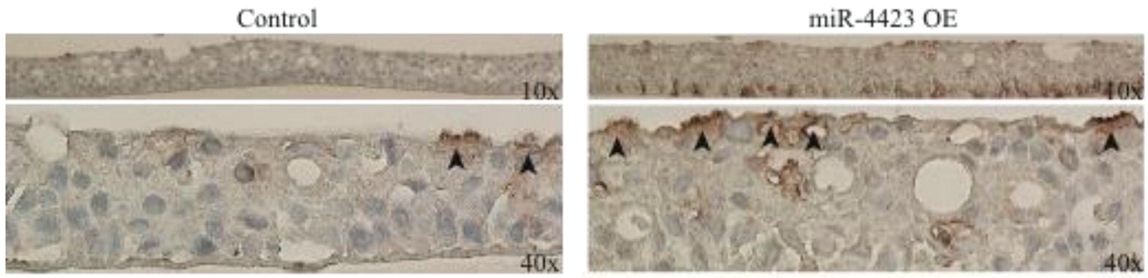

Figure 2.9: Cilia marker β_{IV}-tubulin staining is increased with overexpression of miR-4423. Basal cells infected with lentivirus constitutively overexpressing miR-4423 show more patches of cilia staining for β_{IV}-tubulin at day 11 of the ALI differentiation protocol compared to control. Arrows point to regions of positive staining.

2.2.3 Attempted knockdown of miR-4423 is not sufficient to modulate FOXJ1 expression in ALI cultures

Knock down of both forms of miR-4423 at the same time with a constitutive lentivirus did not affect the gene expression levels of miR-4423, preventing us from measuring the efficiency of the lentivirus in our ALI system. Determining that miR-4423 maybe suppressed and not degraded, we measured the expression of *FOXJ1* to determine if loss of miR-4423 results in a change in the ciliated cell population. No significant change in *FOXJ1* expression was seen with our attempted miR-4423 knockdown (Figure 2.10).

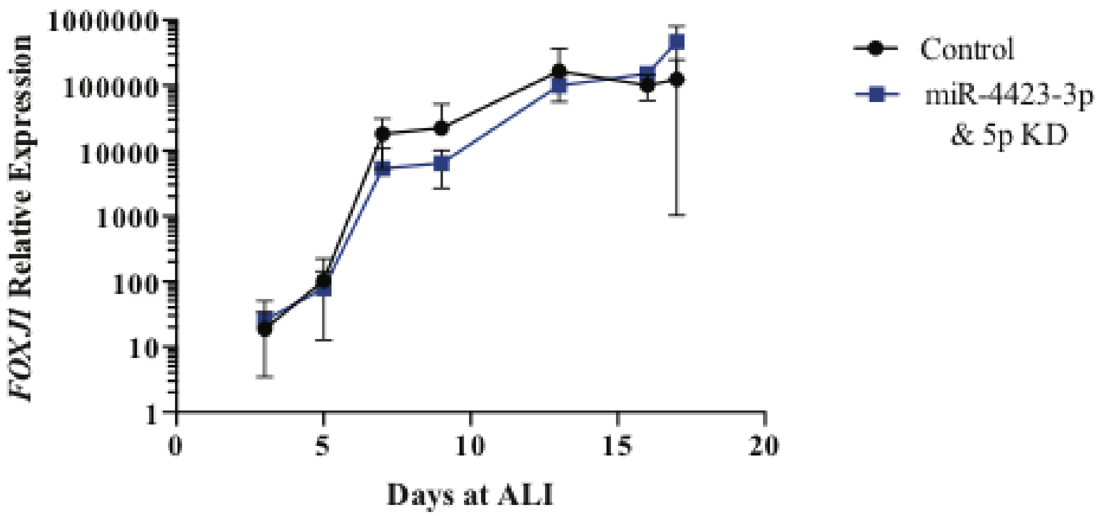

Figure 2.10: Ciliated cell marker *FOXJ1* expression is not changed with miR-4423 knockdown. Basal cells infected with lentiviruses containing antimirs to miR-4423-5p and miR-4423-3p were run on an ALI in two batches. RNA was collected from both batches at multiple time points until day 17 of the differentiation protocol (n=2 to n=6 per time point). No significant difference was observed in the expression of ciliated cell marker *FOXJ1* between the control and knockdown samples.

2.2.4 Expression of miR-4423 is diminished with the presence of cancer

With previous work in the lab showing a field of injury in the airway epithelium that can be used to identify lung cancer in distal regions of the lung[25] we wanted to measure miR-4423 expression in the airway epithelium of patients with and without cancer. Bronchoscopy brushings from patients without cancer had a higher expression of both miR-4423-5p ($p<0.05$) and miR-4423-3p ($p<0.05$; Figure 2.11) than brushings from patients with lung cancer. Similarly, when we measure miR-4423 expression in one patient's samples with increasing severity of disease, starting with laser captured normal airway epithelium, followed by samples of laser captured squamous metaplasia and SCC, we see that miR-4423 expression decreases with increasing disease severity (Figure 2.12).

Moving from the airway to connect to data previously seen with miR-4423's ability to be a tumor suppressor, we measured miR-4423 expression in SCC tumor tissue as well as in the adjacent histologically normal tissue. We see that miR-4423-5p ($p<0.05$) and miR-4423-3p ($p<0.05$) expression is lost with disease (Figure 2.13).

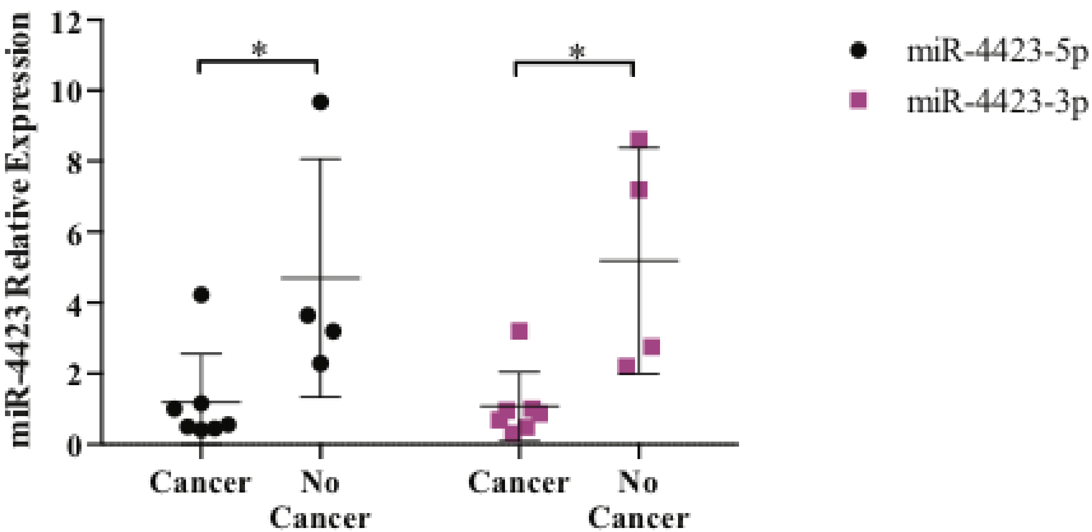

Figure 2.11: Expression of miR-4423 in the histologically normal large airway of patients with lung cancer is lower than that of patients without cancer. Bronchial brushings of histologically normal airway from patients with (n=7) and without (n=4) lung cancer were collected and processed for RNA. Patients with lung cancer have significantly lower expression of miR-4423-5p ($p<0.05$) and miR-4423-3p ($p<0.05$) then patients without lung cancer.

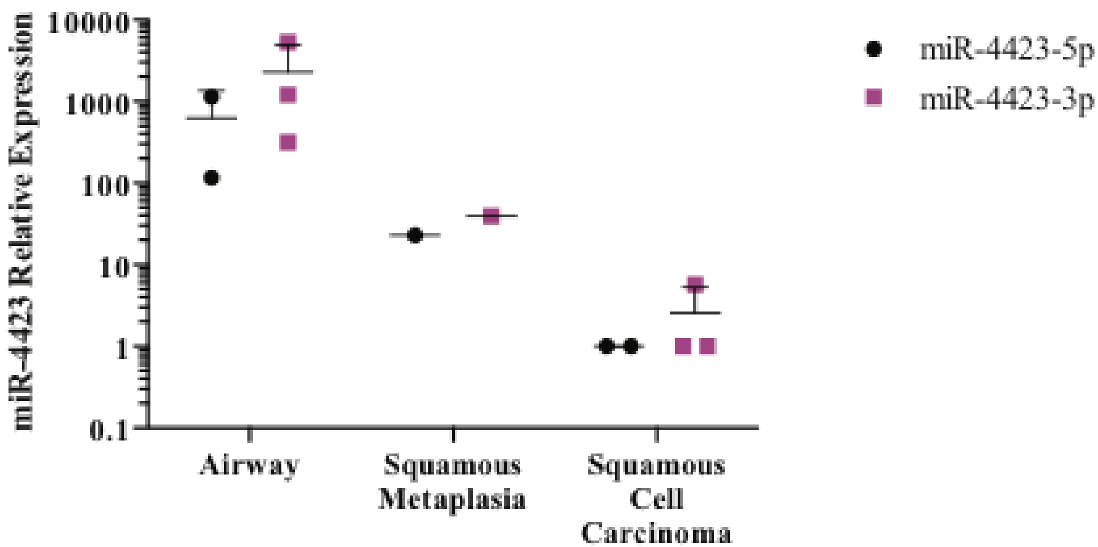

Figure 2.12: Expression of miR-4423 is decreased with increasing disease severity. The expression of miR-4423-5p and miR-4423-3p was measured in laser captured histologically normal airway (n=2-3), laser captured squamous metaplasia (n=1) and SCC (n=2-3) from a resected lung of one patient. Expression of miR-4423 decreases with increasing disease severity: from normal to squamous metaplasia and from squamous metaplasia to SCC.

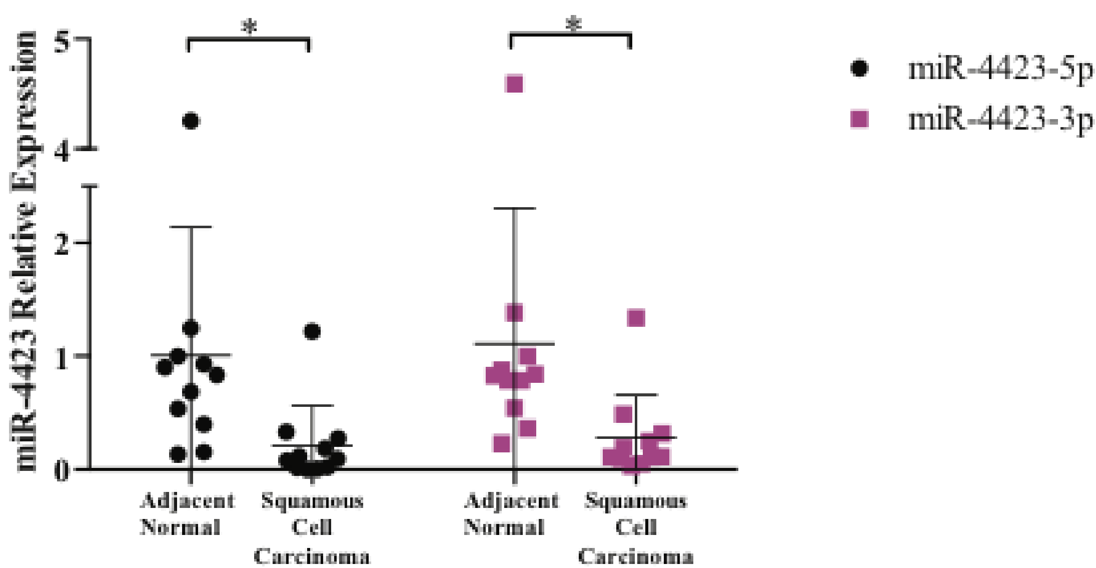

Figure 2.13: Expression of miR-4423 is decreased in tumor tissue compared to adjacent normal tissue. Tumor and adjacent normal tissue from patients with SCC (n=11) were measured for miR-4423 expression. The expression of miR-4423-5p ($p<0.05$) and miR-4423-3p ($p<0.05$) was significantly lower in tumor tissue compared to paired normal tissue.

2.2.5 Potential targets of miR-4423 in tumor cells are not necessarily the same as those in the airway epithelium

In an attempt to identify direct targets of miR-4423, microarray data previously produced in our lab of cancer cell lines sensitive to miR-4423 overexpression (CalU6 and SW900) that were infected with a constitutive miR-4423 lentivirus or an empty vector[96] were mined for genes whose expression was significantly decreased with miR-4423 overexpression. A total of 148 predicted targets of miR-4423 were found to be significantly decreased ($p<0.05$) in the CalU6 cell line and 788 predicted targets of miR-4423 were significant ($p<0.05$) in the SW900 cell line. Between the two cell lines, there were 58 potential targets in common. To narrow down the list of 58 genes, we used *WDR63* as a surrogate marker of miR-4423. A RNA sequencing dataset of ALI inserts over time was mined to identify the correlation of the 58 genes with *WDR63,* looking for genes that are significantly anti-correlated with *WDR63* ($p<0.05$; Figure 2.14). From there, eleven genes were identified as potential direct targets of miR-4423 (Figure 2.15).

Three genes of the eleven were chosen as interesting potential targets of miR-4423 for follow-up studies based on the known biology of the proteins they code for. UBE4B is an ubiquitination factor that monoubiquitinates the tumor suppressor p53 on its own or polyubiquitinates p53 with the help of MDM2[108–110]. KLK10 belongs to a subgroup of serine proteases known as kallikreins. The expression of KLK10 is increased in gastric and colorectal cancer[111,112] while other kallikreins are known to be promote loss of cell-cell junction integrity[113–115]. LRP8 is a member of the low density lipoprotein receptor family that is involved in the Wnt pathway[116] and is a driver of cancer metastasis[117,118]. Known roles in cancer and/or differentiation pathways make *UBE4B*, *KLK10*, and *LRP8* primary targets of interest for direct targets of miR-4423.

Measuring the expression of these three genes in cancer cell lines with miR-4423 overexpression, we see that *UBE4B* expression is reduced in CalU6, SW900 and SKMES1 cell lines (Figure 2.16A). *KLK10* is not endogenously expressed in CalU6, RH2 or H1299 cell lines but in SW900, expression is reduced with miR-4423 overexpression (Figure 2.16B). *LRP8* expression is reduced in CalU6 and SW900 cell lines with miR-4423 overexpression (Figure 2.16C). The two cell lines used to identify potential targets, CalU6 and SW900, show a reduction in the three genes expression with miR-4423 overexpression, but the other cell lines are varied in response.

In ALI samples with miR-4423 constitutively expressed, *UBE4B* (Figure 2.17A), *KLK10* (Figure 2.17B), and *LRP8* (Figure 2.17C) show no significant difference in expression throughout the differentiation process between miR-4423 overexpressing cells and control cultures. While these genes may be direct targets of miR-4423 in tumor cells, they appear to not be affected by miR-4423 expression in the airway epithelium.

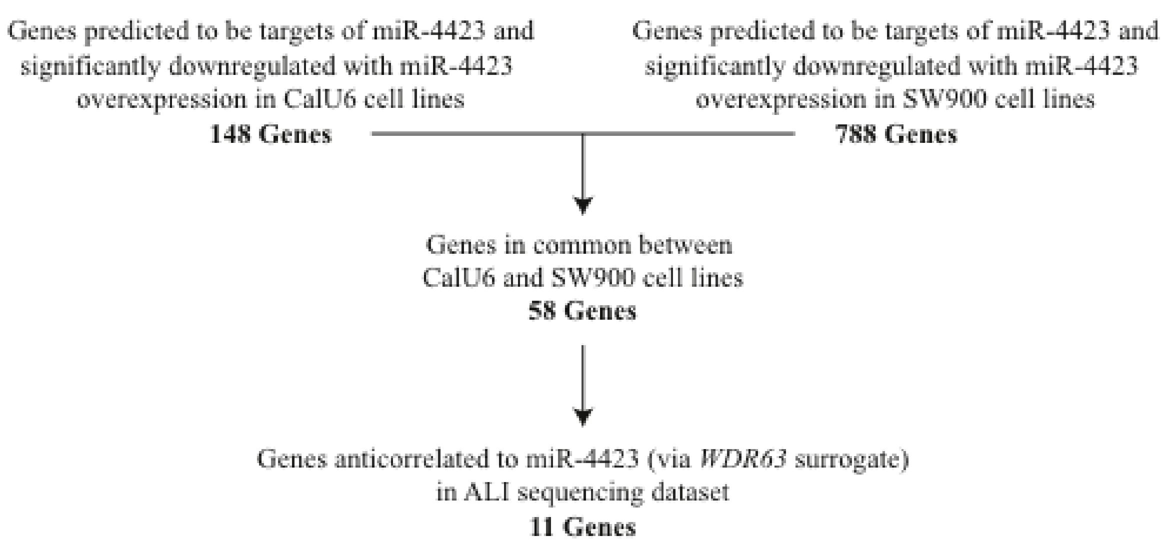

Figure 2.14: Workflow used to identify potential miR-4423 direct targets. All genes identified in bold are significant ($p<0.05$).

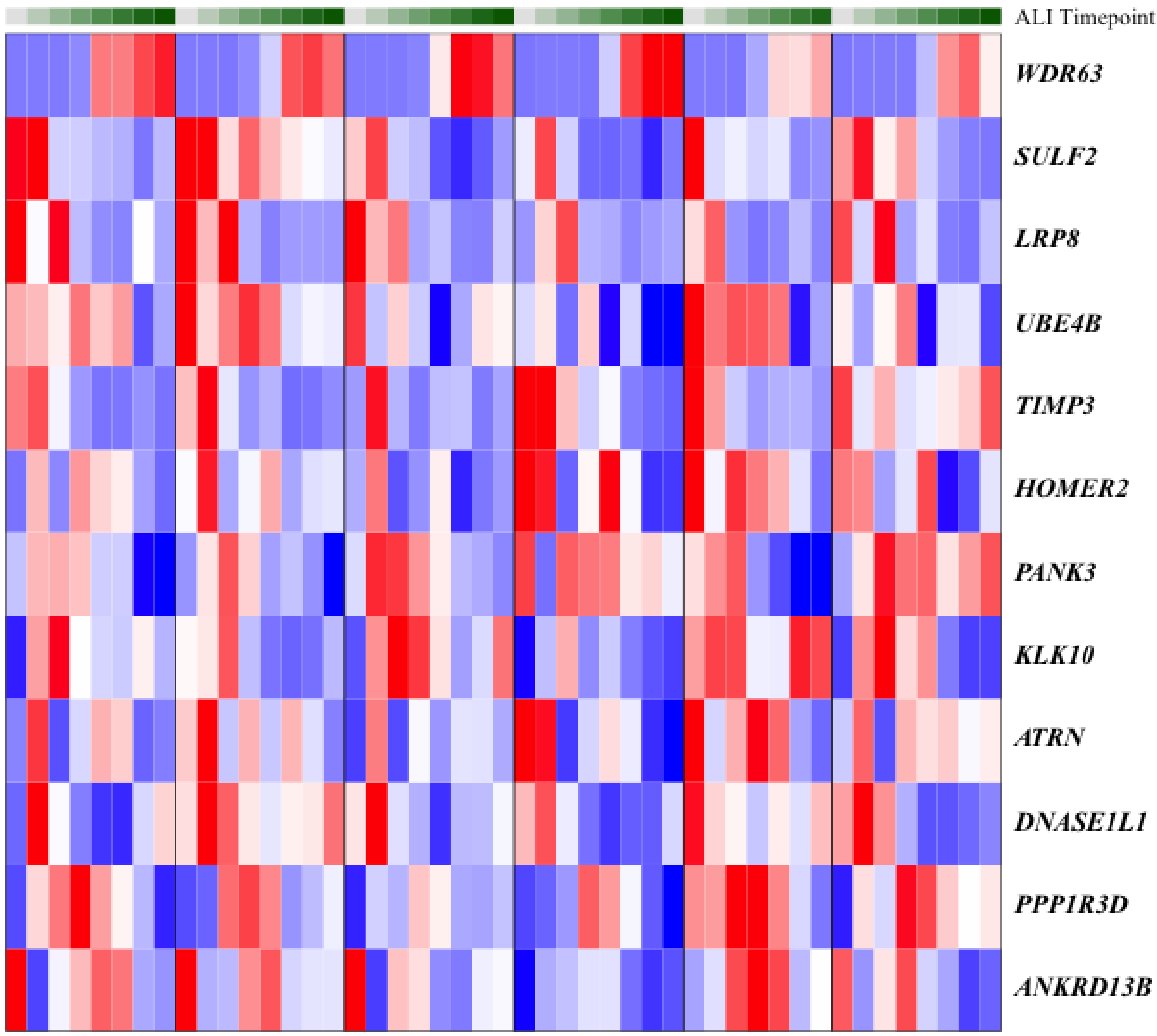

Figure 2.15: Eleven genes are identified in ALI sequencing dataset as anticorrelated to *WDR63* expression. The ALI sequencing dataset is comprised of samples collected at multiple time points throughout the ALI differentiation process with six different donor HBEpCs. Time point during the ALI differentiation process is represented in the color bar at the top of the heatmap with increasing green color with increasing time at an ALI. Using *WDR63* as a surrogate for miR-4423 expression, eleven genes were narrowed down from the 58 genes identified as predicted targets of miR-4423. These eleven genes are significantly anticorrelated to *WDR63* expression at an ALI ($p<0.05$) and are listed in the heatmap based on the strength of their anticorrelation with genes at the top more anticorrelated. Correlations were determined using Pearson correlations.

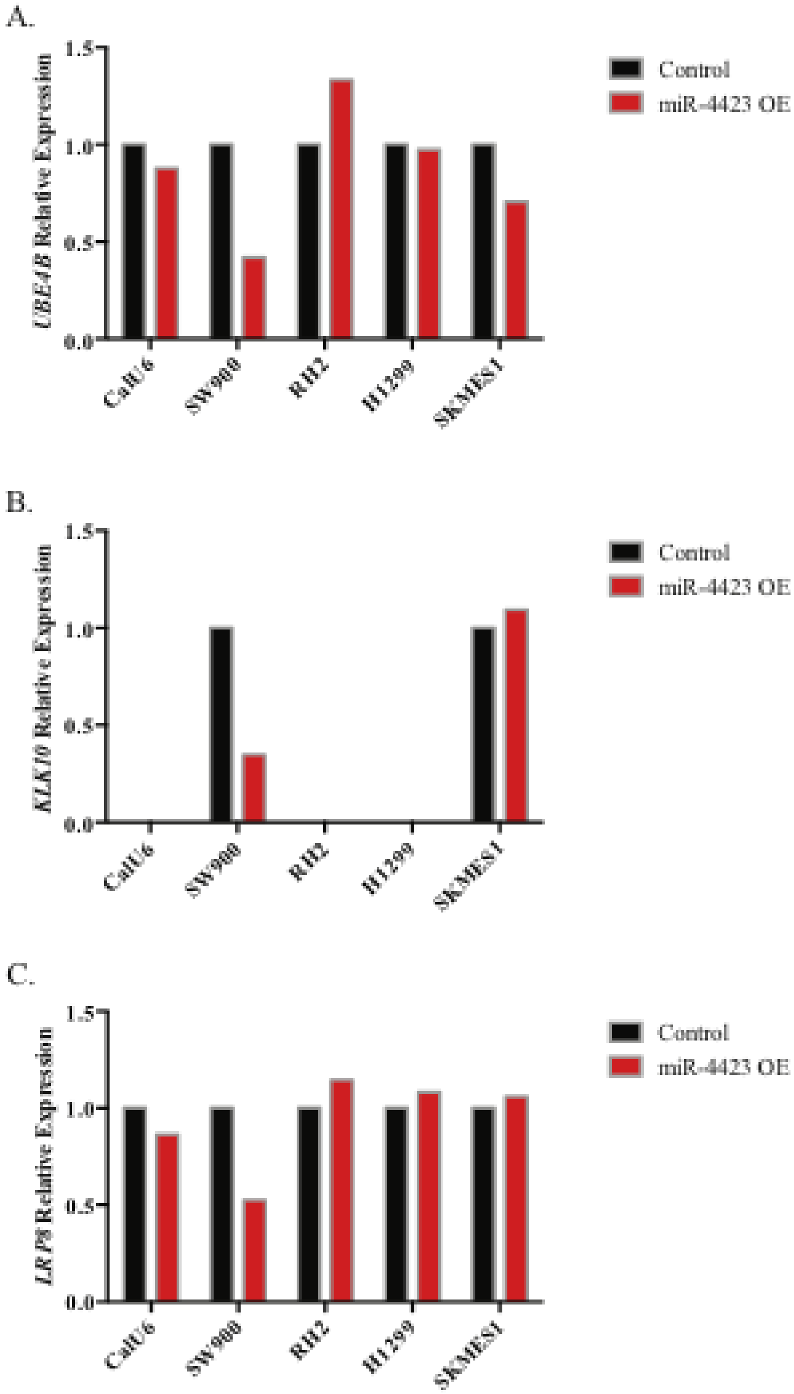

Figure 2.16: Expression of potential direct targets of miR-4423 in cell lines overexpressing miR-4423. Three cell lines that previously showed a decrease in soft agar colonies with miR-4423 overexpression (CalU6, SW900, and RH2) as well as two cell lines that had no change in soft agar (H1299 and SKMES1) were measured for (**A**) *UBE4B*, (**B**) *KLK10*, and (**C**) *LRP8* expression. All three genes show downregulation with miR-4423 overexpression, when the gene is expressed, in the sensitive cell lines CalU6 and SW900.

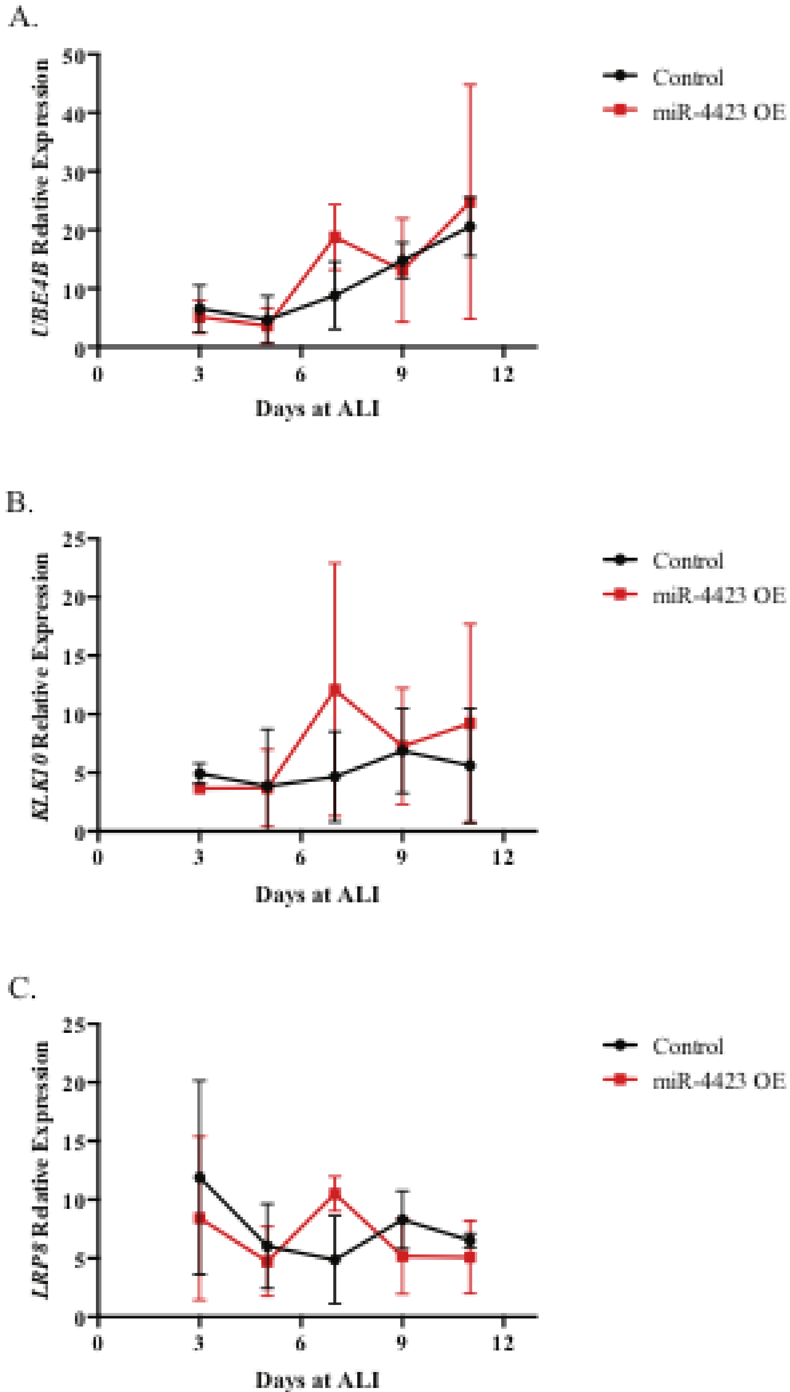

Figure 2.17: Expression of potential direct targets of miR-4423 in ALI cultures overexpressing miR-4423. The expression of the three predicted direct targets of miR-4423 was measured at an ALI overexpressing miR-4423. Basal cells infected with lentivirus constitutively overexpressing miR-4423 show no difference in gene expression of (**A**) *UBE4B*, (**B**) *KLK10*, and (**C**) *LRP8* compared to control ALI. Two ALI batches were normalized together with a total of n=3 to n=4 samples per time point.

2.3 Materials and Methods

Air-Liquid Interface Cultures

Cells used in cultures were obtained from the National Disease Research Interchange (Philadelphia, PA) after a protocol obtained by their IRB was approved. HBEpCs obtained from three donors (two 23-year-old and one 13-year-old nonsmoking Caucasian males) were plated onto collagen coated tissue culture inserts (Millipore Corp., New Bedford, MA) and grown in submerged conditions in BEGM (Lonza, Walkerville, MD) until confluent. Media in the apical chamber was then removed to allow ALI growth. Media in the basolateral chamber was replaced every other day for up to 21 days to allow full differentiation with ALI medium (MatTek Corporation, Ashland, MA) containing insulin (5mg/mL), transferrin (5mg/mL), hydrocortisone (0.5nM), bovine pituitary extract (26mg/mL), EGF (0.6ng/mL), epinephrine (0.5ng/mL), all-trans retinoic acid (50nM), penicillin (100u/mL) and streptomycin (100mg/mL). Cultures were fixed in paraffin sections for evaluation of morphology or placed in RNALater (Ambion, Foster City, CA) until RNA could be isolated.

qRT-PCR

To measure the expression of miR-4453-3p and miR-4423-5p, 10ng of total RNA was used in a custom Taqman MiRNA Assay (Life Technologies, ID #CCBI832, CSFAQ5N, Carlsbad, CA) as per manufacturer's protocol and the results were normalized to RNU44 expression (Life Technologies, Catalog #4427975, ID #001094, Carlsbad, CA). The expression of UBE4B, KLK10, and LRP8 was measured using 500 ng total RNA with the High-Capacity cDNA Reverse Transcription Kit (Life Technologies, Catalog #4368814, Carlsbad, CA) according to the manufacturer's protocol. cDNA product was added to Taqman Fast Universal PCR Master Mix (Life Technologies, Catalog #4352042, Carlsbad, CA) and the appropriate Taqman Gene Expression Assay (Life Technologies, Catalog #4331182, ID #Hs00195897_m1, Hs00173611_m1, Hs00182998_m1, Carlsbad, CA). Data was normalized to the expression of GAPDH (Life Technologies, Catalog #4331182, ID #Hs02758991_g1, Carlsbad, CA) and analyzed using the comparative CT method. The expression of FOXJ1, MUC5AC, MUC5B, CC10, ASCL1, WDR63, and GAPDH were measured by reverse transcribing total RNA using random hexamers (Life Technologies, Carlsbad, CA) and Superscript II reverse transcriptase (Invitrogen, Carlsbad, CA). A total of 20ng of cDNA product was then added to SYBR Green PCR master mix (Life Technologies, Carlsbad, CA) and the relevant primer (IDT, Coralville, IA) for qRT-PCR. Data were normalized to GAPDH expression and analyzed using the comparative CT method.

Primers were designed as follows:

FOXJ1:

For: 5'-GGA GGG GAC GTA AAT CCC TA-3'

Rev: 5'-CCA AAC TTC CAG CTG CTC TC-3'

MUC5AC:

For: 5'-CTG GGA GTC CAG GTC ATG TT-3'

Rev: 5'-TGT TGG CAA ACT TGC TGA AG-3'

MUC5B:

For: 5’-GAC CTG CAA GAA CGA AGA CC-3’

Rev: 5’-AAA GCA CAC GCA CGT TGT AG-3’

CC10:

For: 5’-GCT CAG CTG AAG AAG CTG GT-3’

Rev: 5’-TGA TGC TTT CTC TGG GCT TT-3’

ASCL1:

For: 5’-GAG CAA CTG GGA CCT GAG TC-3’

Rev: 5’CTT TTG CAC ACA AGC TGC AT-3’

WDR63:

For: 5’-GCT GGG AAT AAG CCA CAC AT-3’

Rev: 5’-GAG AGT CCC CCT CTC TGC TT-3’

eGFP:

For: 5’-GAC GTA AAC GGC CAC AAG TT-3’

Rev: 5’-GAA CTT CAG GGT CAG CTT GC-3’

K14:

For: 5’-GGC CTG CTG AGA TCA AAG AC-3’

Rev: 5’-TCC TCA GGT CCT CAA TGG TC-3’

GAPDH

For: 5’-TGC ACC ACC AAC TGC TTA GC-3’

Rev: 5’-GGC ATG GAC TGT GGT CAT GAG-3’

Stable Cell Line Production

The DNA sequence with the miR-4423 precursor and approximately 200bp of flanking region was amplified by PCR using primers (IDT: Forward 5'-CGCGGATCCACACAGCTCACACCATCAGG-3' and Reverse 5'-CGCGGATCCCATGAGACTCTCCCCTGCTC-3', Coralville, IA). The fragment was then ligated into the BamHI restriction site of the phage-CMV-eGFP-W plasmid. Plasmids containing miR-4423 or an empty vector were transfected into 293T cells along with expression vectors containing the packaging proteins Gag-Pol, Rev, Tat and VSV-G. Viral particles in the supernatants were collected and concentrated using ultracentrifugation. Cancer cell lines were grown in 24-well plates and lentiviral infection was performed when cells were 40-50% confluent. HBEpCs (MatTek, Ashland, MA) were cultured on collagen-coated dishes in serum-free BEGM medium (Lonza, Walkersville, MD) to generate stable cells overexpressing miR-4423 or empty vector. Cells were infected with virus particles that were diluted in antibiotic-free BEGM medium with 4ug/mL SureENTRY (Qiagen, Valencia, CA) at an MOI of 20 at 50-60% confluence for 24 hours. To generate HBEpCs with miR-4423 stably knocked-down, primary passage cells were co-infected with both miR-4423-3p and miR-4423-5p miRZip anti-microRNA expression lentivectors or a scramble control (System Biosciences, Mountain View, CA) with an MOI of 20 for each lentivector.

Cell Staining

To prepare ALI samples for staining, EpiAirway tissues (MatTek, AIR-100, Ashland, MA) were fixed in formalin at room temperature overnight followed by embedding in paraffin and sectioning onto slides. Sections were then deparaffinized in Xylene and graded alcohol washes for rehydration. Heat mediated antigen retrieval was performed by incubating slides in 0.01M Sodium Citrate buffer, pH 6.0, for 45 minutes at 100°C. For FOXJ1 staining, slides were blocked in 10% normal goat serum plus 1% BSA in PBS for one hour followed by incubation with anti-FOXJ1 primary antibody (abcam, ab40869, Cambridge, MA) at 1:4000 dilution with 1% BSA in PBS at 4°C overnight. Slides were washed and incubated in Alexa Fluor conjugated secondary antibody (Life Technologies, Carlsbad, CA) at a 1:400 dilution for one hour at room temperature. Sections were imaged using an Olympus FluoView FV1000 confocal microscope. Image analysis and quantification were performed using Nikon NIS Elements software. For β_{IV}-Tubulin staining, Vectastain Elite ABC kit (Vector Laboratories, Burlingame, CA) was used according to the manufacturer's protocol. Sections were incubated in mouse anti-β_{IV}-Tubulin (BioGenex, San Ramon, CA) diluted at 1:1500 with 1% BSA in PBS for one hour at room temperature. Slides were washed and incubated in biotinylated secondary antibody for 30 minutes at room temperature and subsequently incubated for 30 minutes with ABC reagent. Staining was completed with the Vector DAB substrate kit (Vector Laboratories, Burlingame, CA), which was prepared according to the manufacturer's instructions. Slides were incubated in the solution for 5-10 minutes at room temperature followed by a wash with dH_2O and a counterstain with Mayer's Hematoxylin before being coverslipped. Imaging was done with a Nikon Eclipse Ti inverted microscope.

2.4 Discussion

The discovery of the novel primate-specific microRNA miR-4423 led to studies showing miR-4423 is only expressed in mucociliary epithelium and has the ability to inhibit cancer cell growth with forced expression of miR-4423 in some cancer cell lines[96]. Expanding on these studies, we have further characterized miR-4423's role in airway epithelium differentiation and its loss of expression with the progression of lung cancer.

The differentiation of lung airway epithelium is regulated by a number of microRNAs[94,106,107]. For this reason, the study of microRNAs in the airway epithelium is important to better understand the endogenous procedure that basal cells undergo to differentiate into a complete epithelium. By better understanding this process, we can expand our ability to treat many common disease that arise from abnormal airway epithelium, such as asthma, chronic obstructive pulmonary disease and potentially even lung cancer[4,27,103]. The work done in this study characterizes a newly discovered microRNA in the airway epithelium by utilizing an ALI culture system. We show that expression of miR-4423 is increased during the basal cell differentiation process, with little to no expression in basal cells themselves, and that the endogenous expression of miR-4423 closely correlates with expression of ciliated cell markers *FOXJ1* and *WDR63*. Additionally, when miR-4423 is stably overexpressed in basal cells, we see an increase in ciliated cells during the differentiation process. These results indicate miR-4423 plays a role in promoting ciliogenesis. After attempting to knockdown miR-4423 we see no change in gene expression of *FOXJ1* and no phenotype change in the airway epithelium, implying that knocking down miR-4423 is not sufficient to alter ciliogenesis in the airway epithelium. With miR-4423 being a primate-specific microRNA, it's possible that it plays a redundant role and that there are other family members that can compensate for loss of miR-4423. Indeed, the miR-34/449 family has members whose seed sequences are close to those of miR-4423[96] and this family is very well characterized as being evolutionarily conserved and important in ciliogenesis in the airway epithelium[94,107,119].

Expanding on studies showing miR-4423's potential role as a tumor suppressor in lung cancer[96], as well as previous work showing a field of injury in the airway of patients with lung cancer[25], we measured miR-4423 expression in histologically normal airway epithelium of patients with and without lung cancer. The expression of both forms of miR-4423 is diminished in the airway of patients with lung cancer, implying that miR-4423's loss of expression indicated a microRNA field of injury for cancer. Moving out of the airway and into premalignancy, we measured miR-4423 expression in laser captured sections of one patient's normal airway, squamous metaplasia and SCC tissue. While the sample numbers are small and only representative of one patient, it appears that the expression of miR-4423 is gradually lost with increasing disease. When looking at miR-4423 expression in tumor compared to adjacent histologically normal tissue, there is a significant loss of miR-4423 expression in SCC tissue compared to normal tissue. All of these results indicate a true loss of miR-4423 expression with the development of lung SCC and reinforce the concept of miR-4423 as a tumor suppressor.

With miR-4423 playing a role in promoting ciliogenesis and acting as a tumor suppresser when there is forced expression in the tumor cells, it stands to reason that the mechanism by which miR-4423 promotes differentiation is the same as the mechanism in which it inhibits cell growth in tumors. With the cell of origin for squamous cell carcinoma not believed to be ciliated cells, we can postulate that loss of miR-4423 in ciliated cells is not a driver of lung cancer, but a consequence. Preliminary analysis of predicted targets of miR-4423 in our datasets, has not identified a shared mechanism for ciliated cell differentiation and cancer inhibition with miR-4423 expression. Three genes that may play a role in miR-4423's ability to hinder tumor growth do not appear to be altered with miR-4423 overexpression at an ALI. A better understanding of miR-4423's mechanism of action in tumor and airway phenotypes is necessary for further research into miR-4423's potential as a cancer therapeutic.

This study is not without limitations. All ALI cultures have a large batch affect, making it hard to combine experiments. In addition, while we can measure the overexpression of miR-4423 in cell culture, the knockdown of miR-4423 does not degrade the microRNA, meaning that by qRT-PCR we are unable to measure the knockdown efficiency and can only rely on being able to measure transfection efficiency of the lentiviruses via GFP expression in the cells. Future work will need to be done to uncover the mechanism of action for miR-4423 in ciliogenesis as well as its role in suppressing tumor growth.

2.5 Conclusions

The research put forth in this study clearly shows a role for miR-4423 in airway epithelial differentiation. Moreover, we show that miR-4423 expression is gradually lost with increasing disease severity in relation to SCC. These findings as well as work previously done in our lab showing miR-4423's ability to suppress cancer growth[96], indicate a potential for miR-4423 to be utilized as a therapeutic for SCC. Further research is needed to determine if miR-4423 intervention can reverse early events in carcinogenesis, such as premalignant lesions. The mechanism of action for miR-4423 in regards to its ability to promote ciliogenesis as well as its ability to suppress tumor growth needs to be better understood, whether there is a common mechanism for both phenotypes or individual pathways that are affected based on the biological context, remains to be seen. Finally, the characterization of this microRNA in the airway epithelium highlights the need to better understand the role of microRNAs in normal biological processes so that we can better understand their part in disease progression and how they can be utilized as therapeutics.

CHAPTER THREE

MicroRNA 424 is a Potential OncomiR in Never Smoker Lung Adenocarcinoma that Regulates Cell Migration

3.1 Introduction

While the leading cause of lung adenocarcinoma is cigarette smoke exposure, 10-15% of lung adenocarcinoma cases in the US occur in patients who have never smoked[120]. Lung cancer in never smoker patients is the seventh leading cause of all cancer mortality and the top cancer related cause of death of never smokers in the US[39,97]. Outside the US, never smoker lung cancer rates are highest in China and appear to be on the rise in a number of countries including: Japan, Sweden, Italy and the United Kingdom[121]. The incidence of lung cancer amongst never smokers in the United States continues to rise annually[122].

Recent studies have indicated that lung cancer patients who are either ever smokers (>100 cigarettes in their lifetime) or never smokers harbor distinct profiles of somatic mutations, and that these two groups elicit disparate responses to targeted therapy[123,124]. For example, *EGFR* mutations are present in 45% of never-smoker lung adenocarcinomas but only 7% of ever-smoker lung adenocarcinomas, and can be exploited for targeted therapy with EGFR-TKI (Tyrosine Kinase inhibitors)[40]. Although the use of EGFR-TKI offers therapeutic benefits to some patients with late stage adenocarcinoma, many develop resistance to EGFR-TKI, rendering the use of this therapeutic limited[125]. This supports the need for further molecular characterization of never smoker lung cancer to develop novel therapeutics. In addition, a number of molecular abnormalities including *TP53* mutations, *KRAS* mutations, chromosomal aberrations, and unique gene expression and methylation profiles, have been identified to vary between the lung tumors of never smokers and ever smokers suggesting that there exist unique molecular drivers that require further characterization[126–128].

Previously we performed large and small RNA sequencing on paired adenocarcinoma and adjacent histologically normal tissue from ever and never smokers. Linear modeling revealed 120 large RNA (Figure 3.1A) and 15 microRNA (Figure 3.1B) uniquely changing in never smokers.

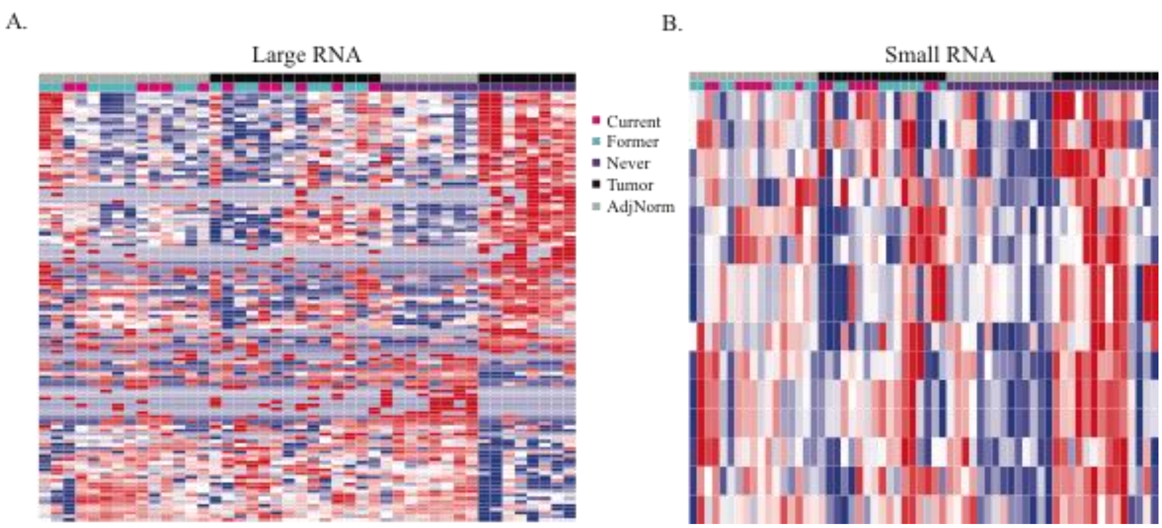

Figure 3.1: Large and small RNA sequencing identifies never smoker tumor-specific transcriptomic perturbations. (**A**) Semi-supervised heatmap of 120 genes whose expression levels vary uniquely between tumor and adjacent-normal samples in never smokers. Differential expression analysis was conducted using the same linear mixed effect models and filters as implemented for the microRNA data. Each row represents one of 120 genes, and each column represents one of 44 tumor and adjacent normal samples collected from 22 subjects. Blue represents lower relative expression, and red represents higher relative expression. (**B**) Semi-supervised heatmap of 15 microRNAs whose expression levels vary uniquely between tumor and adjacent-normal samples in never smokers. Using the same differential expression strategy as for large RNA, linear modeling identified 15 microRNAs that appear to be uniquely modulated in the never smoker tumor.

To gain insight into the regulation of never smoker lung cancer, mirConnX was used to build a regulatory network[129]. This network contains 7 microRNAs (miR-424, miR34a, miR-500, miR-1260, miR-652, miR-324, miR-146b) connected to 519 protein coding genes, 18 of which are transcription factors, for a total of 591 interactions (Figure 3.2).

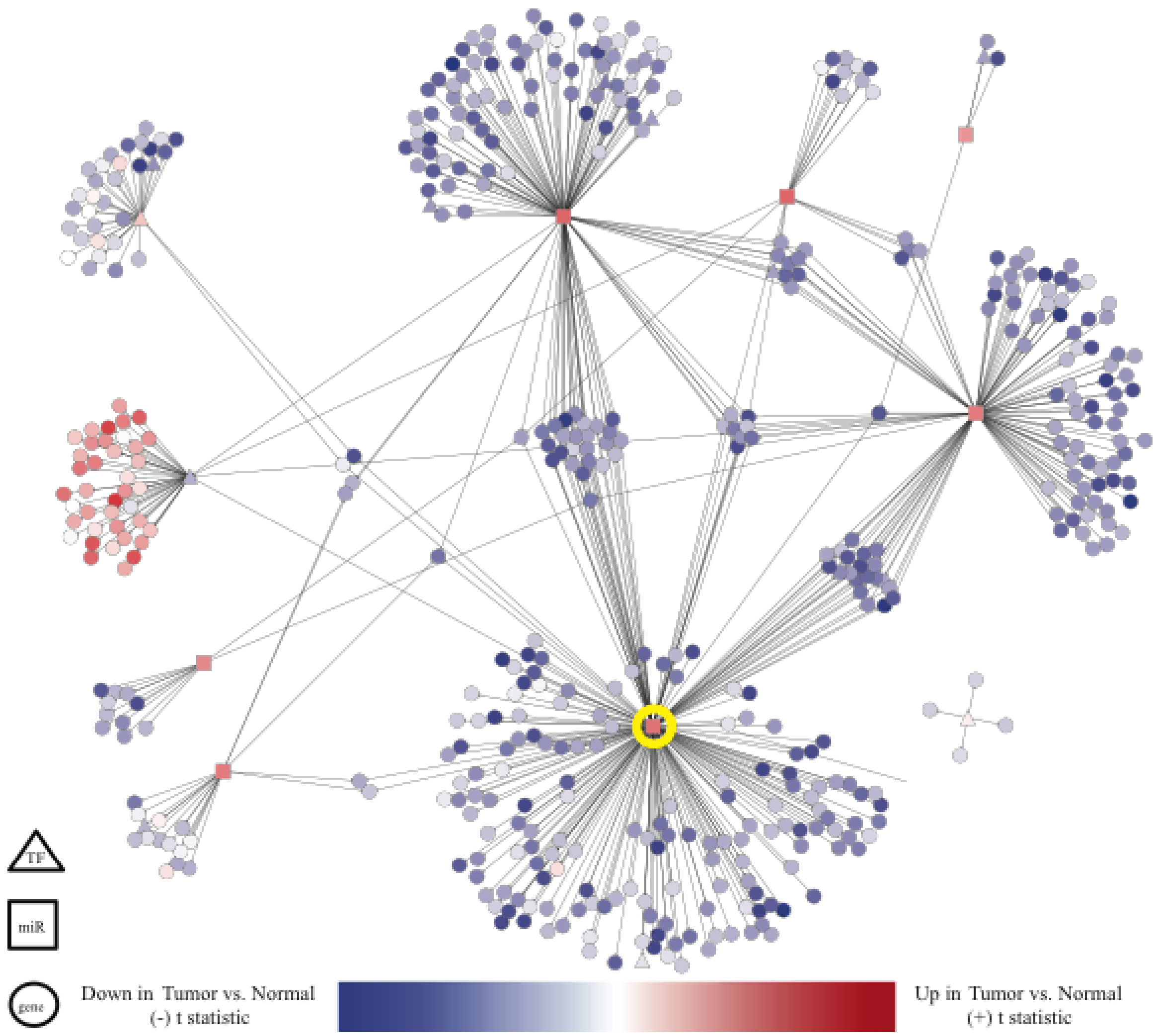

Figure 3.2: Tumor-associated microRNA-mRNA network constructed from never smoker sequencing results implicates miR-424-5p as a regulatory hub. The 15 microRNAs that were predicted to be associated with never smoking adenocarcinoma were integrated with gene expression data from matching samples to construct a microRNA-mRNA regulatory network (mirConnX). After overlaying the filtered expression data with a prior network of putative target predictions, we computed a network comprised of 591 interactions, which includes 7 microRNAs (squares) exhibiting suppressive interactions with 501 protein-coding genes (circles) and 18 transcription factors (triangles). One microRNA, miR-424-5p (circled in yellow) was identified as a major regulatory hub and selected for further analysis. Node colors reflect the relative high (red) or low (blue) T/N expression of each mRNA or microRNA as calculated by the T statistic generated from a Student's t test.

Through the combination of RNA sequencing and integrative transcriptomic network analysis, potential microRNA regulators of never smoker lung adenocarcinoma are revealed. Further functional validation and molecular biology are needed to parse down the list of seven microRNAs and identify the phenotypes affected by their overexpression in tumor tissue. Extending the studies presented, we sought to explore the microRNA regulation of never smoker lung adenocarcinoma.

3.2 Results

3.2.1 Identification of miR-424 as a potential regulatory element of never smoker lung adenocarcinoma

One microRNA, miR-424, was highly connected in this network (with 218 connections) and inconsistently expressed in the tumors of ever smokers compared to paired adjacent normal tissue (Figure 3.3A), while it was expressed higher in tumor tissue compared to paired adjacent normal in individual never-smoker patients (Figure 3.3B).

The 218 genes connected to miR-424 in the mirConnX network (Table 3.1) were further analyzed for function enrichment using a pathway analysis tool called Enrichr[130]. Many cancer-related pathways were identified as significantly represented in our list of genes, including: regulation of cell adhesion ($p<0.000005$), regulation of epithelial cell migration ($p<0.0005$), and PI3K/AKT signaling in cancer ($p<0.005$; Table 3.2).

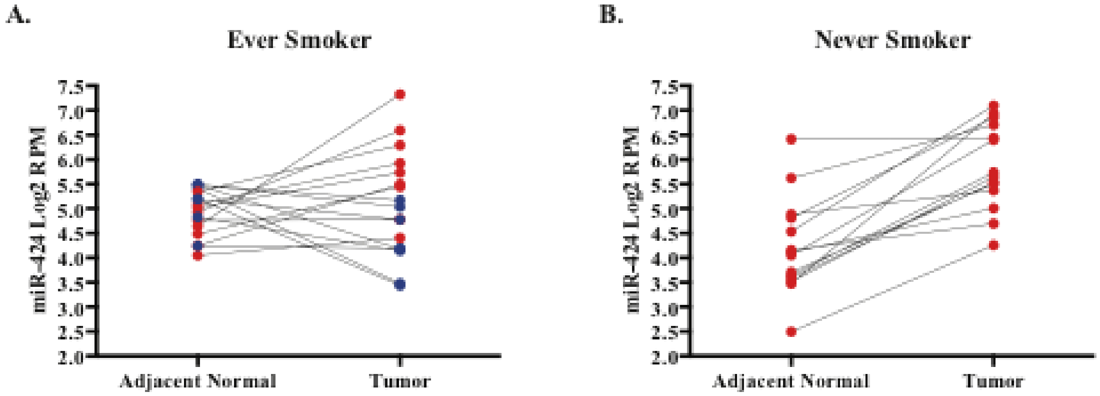

Figure 3.3: Dynamics of RNA sequencing miR-424 expression in individual patients. The paired tumor and adjacent normal RNA sequencing expressions are represented with red indicating higher expression in the tumor than adjacent-normal and blue indicating lower expression in the tumor than adjacent-normal in (**A**) ever smokers with lung adenocarcinoma (n=17) and (**B**) never smokers with lung adenocarcinoma (n=14).

ABCG4	C20orf46	DNAJB4	FRY	LRRC7	PDPN	RSPO2	ST8SIA1
ABHD2	C4orf32	DOCK4	FZD10	LRRN3	PID1	RSPO3	SV2B
ACSL4	CA2	DPT	FZD4	MAMDC2	PIK3R1	RUNX1T1	SVEP1
ADAMTS3	CACNA2D3	DSEL	GHR	MAP2	PIP5K1B	SCN4B	SYN2
ADAMTS5	CADM1	DTNA	GJA5	MASP1	PLAG1	SEMA3C	TACC1
ADAMTSL3	CAPN6	DUSP26	GNG2	MEOX2	PLCXD3	SEMA3D	TDGF1
ADRB2	CCND1	DYNC1I1	GPR88	MFAP5	PLEKHC1	SEMA6D	TGFBR3
AFF2	CD34	EDA	GRIA3	MN1	PLSCR4	SESN1	TIMP2
AHNAK	CD8B	EDNRB	HAS2	MOBKL2B	PMP22	SGCD	TIMP3
AKAP2	CDC42EP2	EFNB1	HEPH	MS4A2	PODXL	SH2D3C	TLR4
AKT3	CDH11	EFNB2	HLF	MS4A7	PPAP2B	SH3GL2	TMEM100
AMOTL1	CDK6	EIF5A2	HOXA3	MSR1	PPARGC1A	SH3TC2	TMEM47
ANK2	CHRDL1	EMCN	HPSE2	MYCT1	PPP1R14C	SLC16A1	TMEM86A
AQP4	CITED2	EPHA7	HS3ST2	MYLK	PRKG1	SLC24A3	TMOD1
ARHGAP18	CLCN1	ETS1	HTR2A	NEBL	PRRX1	SLC25A24	TNFAIP8L3
ARHGAP20	CLEC12B	FAM70A	IRS2	NEO1	PTGER3	SLC2A3	TRAM1
ARL10	CNN1	FAM89A	KCNAB1	NLGN1	PTGER4	SLC5A1	TSGA13
BACH2	COL12A1	FAT3	KCNIP1	NTRK2	PTPRD	SLITRK2	TSPAN7
BDNF	CREB5	FGF12	KCNT2	NUAK1	QKI	SLN	TUSC3
BMX	CRIM1	FGF2	KDR	ODZ2	RAB8B	SMOC1	VGLL3
BSND	CYBB	FGF7	KIAA1045	OSR1	RASGEF1B	SNAI2	VLDLR
BTG2	CYBRD1	FGFR2	LCP1	PCDH17	RASL12	SOBP	VWF
BVES	CYGB	FIGF	LMOD1	PCDH9	RECK	SPSB4	WNT3A
C10orf54	DACH1	FKBP5	LPHN2	PCDHA10	RELN	SPTBN1	WNT7A
C11orf9	DGKG	FOXP2	LRCH1	PCDHA12	RGS9BP	SRGN	ZCCHC5
C12orf64	DIXDC1	FRAS1	LRP2	PDE3B	RIMS3	SRPX	ZNF423
C1QB	DLL1	FRMD6	LRRC55	PDK4	RNF125	ST6GALNAC3	ZNF536
C1QL3	DLL4						

Table 3.1: Genes that are connected to miR-424-5p node in the mirConnX network.

Tool	Pathway	P-value
GO	Regulation of cell adhesion	<0.000005
GO	Cell migration involved in sprouting angiogenesis	<0.0001
GO	Canonical Wnt signaling pathway	<0.0001
GO	Regulation of epithelial cell migration	<0.0005
WikiPathways	Focal adhesion	<0.05
Reactome	PI3K/AKT signaling in cancer	<0.005
KEGG	Non small cell lung cancer	<0.05

Table 3.2: Enrichr analysis of genes predicted to be targets of miR-424-5p in the mirConnX network.

3.2.2 Validation of miR-424 expression

The potential regulatory hub miR-424-5p and transcription factor, *FOXP2,* were identified from the RNA sequencing data and constructed networks as being uniquely altered in the tumor versus adjacent-normal tissue of never smokers as compared to ever smokers. The behavior of these genes was subsequently validated by qRT-PCR using a composite of adenocarcinoma tumor and adjacent-normal samples from an independent sample set and remaining RNA from the original sample set analyzed together (10 current, 22 former, 20 never; Table 3.3). MiR-424 fold change in paired tumor and adjacent normal samples is significantly different between never and ever smokers in the RNA sequencing data ($p<0.005$; Figure 3.4A). The difference seen is recapitulated by qRT-PCR ($p<0.05$; Figure 3.4B). Similarly, *FOXP2* is downregulated in the tumors of never smokers compared to ever smokers in the RNA sequencing data ($p<0.05$; Figure 3.5A) and validates by qRT-PCR ($p<0.05$; Figure 3.5B).

EGFR mutational status of the independent sample set (Table 3.4) was known and miR-424-5p expression shows no significant difference in ever smokers ($p=0.74$) and never smokers ($p=0.87$) based on mutation status (Figure 3.6).

Previous work in the lab shows a field of injury in the airway epithelium that can be used to identify lung cancer in distal regions of the lung[25]. We wanted to measure miR-424 expression in the airway epithelium to explore the possibility of a field in never smokers. The expression of miR-424-5p was measured by qRT-PCR in cytologically normal small airway samples from patients with lung adenocarcinoma (5 current, 14 former, 14 never; Table 3.4). We observed higher expression of miR-424 in the small airways of never-smokers compared to ever-smokers (p<0.05; Figure 3.7).

	Current Smokers	Former Smokers	Never Smokers
Tumor/Adjacent Normal Pairs	10	22	20
Gender	3 Male 7 Female	10 Male 12 Female	8 Male 12 Female
Age	66.7 +/- 7.8	68.0 +/- 7.9	64.8 +/- 10.2

Table 3.3: Cohort of clinical samples used for qRT-PCR validations.

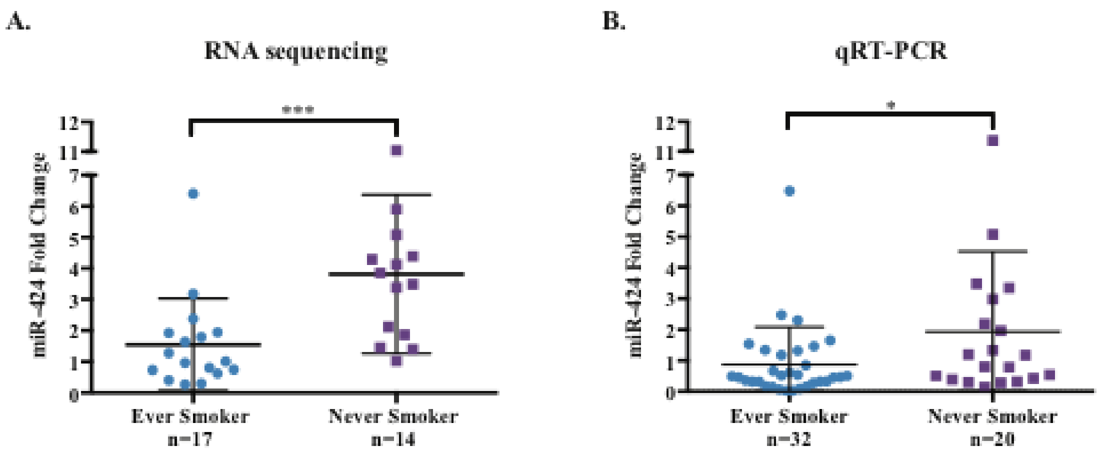

Figure 3.4: MiR-424-5p expression by microRNA sequencing is validated by qRT-PCR. (**A**) Comparison of tumor and adjacent normal expression levels for miR-424-5p as detected by microRNA sequencing. log2 FPKM of miR-424 in the paired tumor and adjacent-normal microRNA sequencing data was used to calculate fold change (FC). Never smoker paired samples demonstrate significantly higher expression of miR-424 in tumor versus adjacent normal than the ever smoker paired samples (never smoker tumor vs. adjacent normal p-value<0.0005). (**B**) qRT-PCR results indicating miR-424 expression levels in tumor versus adjacent normal samples from a cohort of previously sequenced and independent samples. miR-424-5p expression levels in tumor versus adjacent normal samples from never smokers (n=20) demonstrate significant two-fold increase (p<0.05) compared to levels in tumor versus adjacent normal of ever smokers (n=32).

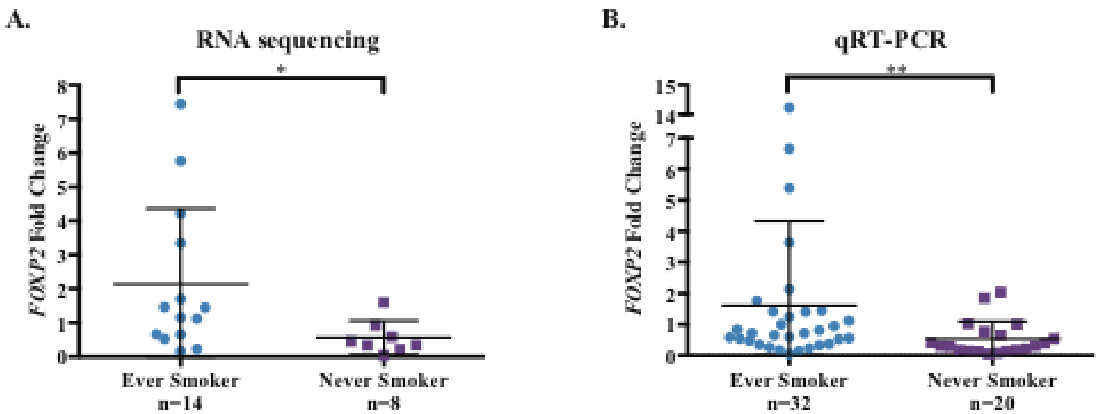

Figure 3.5: *FOXP2* expression by large RNA sequencing is validated by qRT-PCR. (**A**) Expression of *FOXP2* in the paired tumor and adjacent-normal large RNA sequencing data was used to calculate FC. Never smoker paired samples show significantly lower ($p<0.05$) expression of *FOXP2* in tumor vs. adjacent-normal then the ever smoker paired samples. (**B**) qRT-PCR of ever (n=32) and never (n=20) paired tumor and adjacent-normal samples show a 2-fold decrease in *FOXP2* expression levels in the tumor of never smokers compared to their adjacent-normal, a significant decrease compared to ever smokers ($p<0.005$).

	Current Smokers	Former Smokers	Never Smokers
Tumor/Adjacent Normal Pairs	5	14	14
Gender	2 Male 3 Female	5 Male 9 Female	6 Male 8 Female
Age	67 +/- 7	71 +/- 9	64 +/- 10

Table 3.4: Cohort of clinical samples used for *EGFR* and small airways qRT-PCR studies.

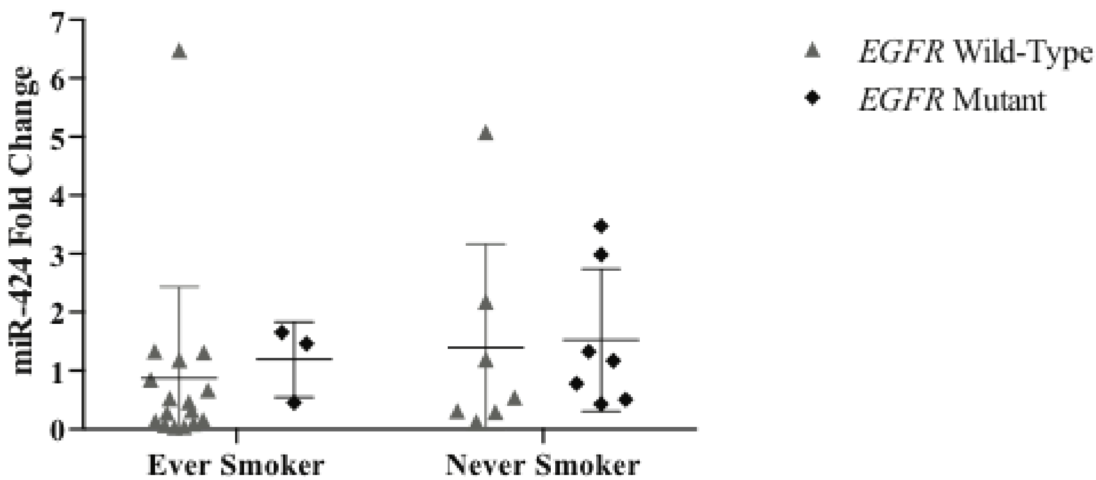

Figure 3.6: MiR-424-5p expression is not affected by *EGFR* status. *EGFR* status of samples from MD Anderson Cancer Center (n=5 currents, 14 formers, 14 never smokers) was identified and miR-424 FC was plotted based on status. No significant differences were observed.

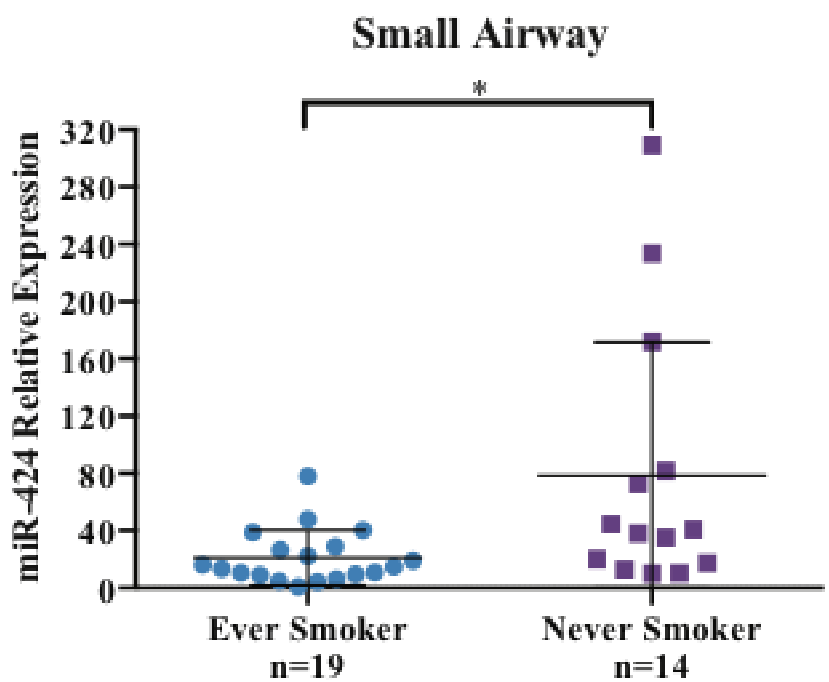

Figure 3.7: Expression of miR-424-5p is higher in the small airway of never smokers with lung adenocarcinoma compared to ever smokers. miR-424-5p was measured in small airways samples from the same patients that paired tumor and adjacent normal were obtained from at MD Anderson Cancer Center (n=5 currents, 14 formers, 14 never smokers). miR-424-5p expression was highly expressed in the small airways of never smokers compared of those of ever smokers ($p<0.05$).

3.2.3 Pathway enrichment of perturbing miR-424 in never smoker lung cancer cell line

To characterize the transcriptomic changes induced by miR-424-5p perturbation, we profiled gene expression in the never-smoker lung adenocarcinoma cell line H2085 transfected with either an inhibitor of miR-424-5p (n=3) or a control inhibitor (n=3) using microarrays. All profiled genes were organized into a ranked list by t-statistic and tested for enrichment of the mirConnX-predicted miR-424 targets from the regulatory network. GSEA revealed statistically significant positive enrichment of this set of genes ($p<0.05$), confirming mirConnX's ability to predict association between a microRNA and its targets. We probed the leading edge of this analysis for enriched pathways. The leading edge results indicated that miR-424-5p might be regulating migration, angiogenesis, and cell adhesion in our never-smoker lung adenocarcinoma cell line (Table 3.5).

Tool	Pathway	P-value
GO	Cell migration involved in sprouting angiogenesis	<0.00005
GO	Endothelial cell migration	<0.005
GO	Regulation of cell adhesion	<0.005
GO	Regulation of ERK1 and ERK2 cascade	<0.005
GO	Epithelial cell migration	<0.005
WikiPathways	Regulation of actin cytoskeleton	0.005
WikiPathways	Angiogenesis	<0.0005
PPI Hub Protein	EGFR	<0.0005

Table 3.5: Enrichr analysis of genes in the mirConnX/microarray GSEA leading edge.

3.2.4 Potential targets of miR-424

The identification of potential targets of miR-424 required numerous steps. To begin with, a list of predicted targets of miR-424 from TargetScan[131] were compared to the 882 genes significantly upregulated with miR-424 knockdown as measured in our microarray analysis (Figure 3.8). From there, the 154 genes identified were compared to the genes interacting with miR-424 in the mirConnX map. A total of ten genes fit all the filtering criteria (Table 3.6). Of interest for possible follow-up studies, *DNAJB4* is reported to post transcriptionally influence E-cadherin localization and stability, making it a potential tumor suppressor[132]. *TMEM100* is downregulated in lung adenocarcinoma and in-vitro studies show that overexpression of *TMEM100* inhibits colony formation in lung cancer cell lines[133]. Additionally, *TMEM100* is published as a potential tumor suppressor in hepatocellular carcinoma[134].

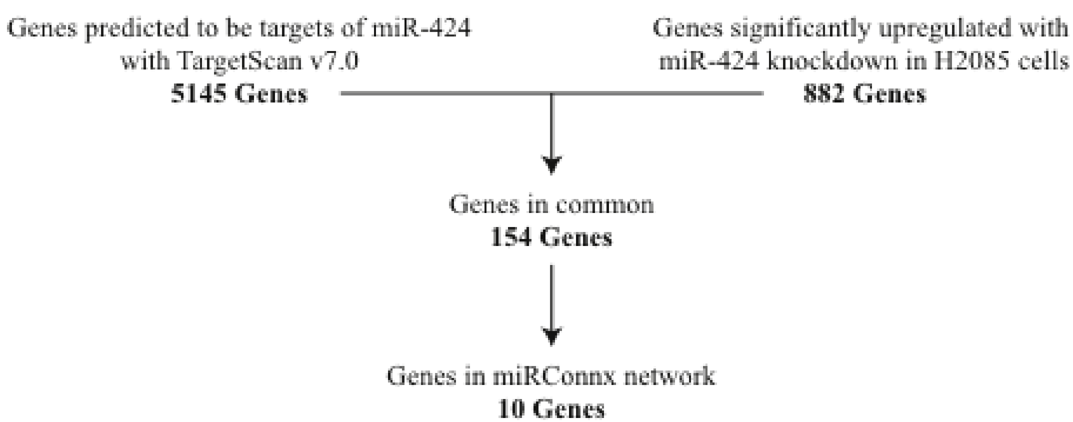

Figure 3.8: Filtering criteria for predicted miR-424 direct target identification.

CDK6	EMCN	HAS2	LCP1	ST8SIA1
DNAJB4	FGFR2	HOXA3	SGCD	TMEM100

Table 3.6: Predicted direct targets of miR-424 identified through a multiple step filtering criteria.

3.2.5 MiR-424 modulates migration in cancer cell lines

Following up on the pathway enrichment data implicating miR-424 regulates migration related pathways, we measured migration *in vitro* via scratch assays in lung adenocarcinoma cell lines developed from both smokers and never smokers. After knocking down miR-424, lung adenocarcinoma cell lines derived from ever smokers (Figure 3.9A) and never smokers (Figure 3.9B) were screened for a migration phenotype. The never smoker lung adenocarcinoma cell line H2085 was identified as having the strongest migration phenotype. In the H2085 cell line, efficient miR-424 knockdown results in a reduction of overall cell migration (Figure 3.10A, Figure 3.11A). This result is reproducible, with the averages of three separate experiments showing a reduction in overall cell migration by 10-20% (Figure 3.11B). Interestingly, overexpression of miR-424 in a cell line with limited migration capacity, SK-LU-1, results in an increase of migration (Figure 3.10B, Figure 3.11C) that is consistent (Figure 3.11D).

To ensure what we see by scratch assay is migration and not a result of proliferation, the scratch assays were stained for F-Actin at the time of scratch induction and at 24 hours after scratch induction. We see a decrease in the number of H2085 cells with a strong leading edge in cells transfected with anti-miR-424 (Figure 3.12A) and an increase in the number of SK-LU-1 cells with a strong leading edge with transfection of miR-424 mimic (Figure 3.12B). The lack of a proliferation phenotype affecting the scratch assay results is additionally supported by no change in cell cycle with miR-424 knockdown (Figure 3.13A, Figure 3.13B) and no change in the number of cells staining for H3Sp10, a marker of mitosis (Figure 3.14A). Cells staining positive for H3Sp10 are decreased, though not significantly, with a reduced serum environment (Figure 3.14B) but remain unchanged between cells transfected with control or anti-miR-424 in a serum rich media (Figure 3.14C). Growth in a low serum environment, transfected with control or anti-miR-424, and scratched also results in no change in H3Sp10 staining (Figure 3.14D).

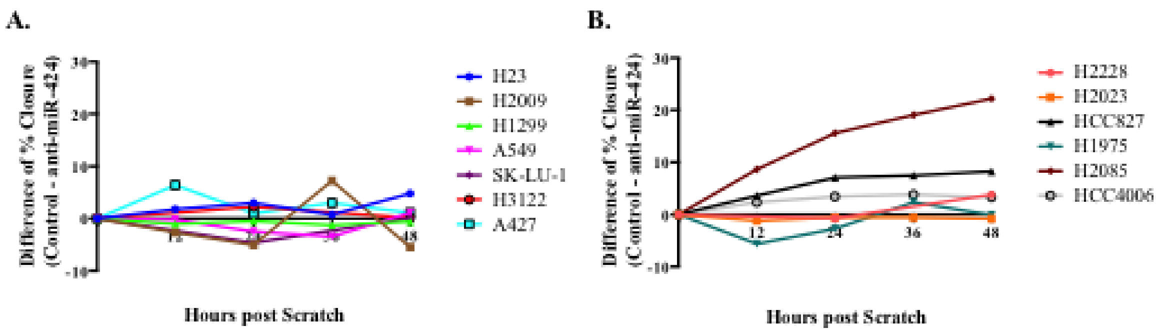

Figure 3.9: Initial screening of lung adenocarcinoma cell lines identifies H2085 as having the strongest migration phenotype. Lung adenocarcinoma cell lines derived from (**A**) ever and (**B**) never smokers were transfected with anti-miR-424 or control and measured for their capacity to close the induced scratch over 48 hours. Data shown is the difference in the % closure of the control – anti-miR-424 samples, therefore a higher % closure indicates a reduction in migration with the knockdown.

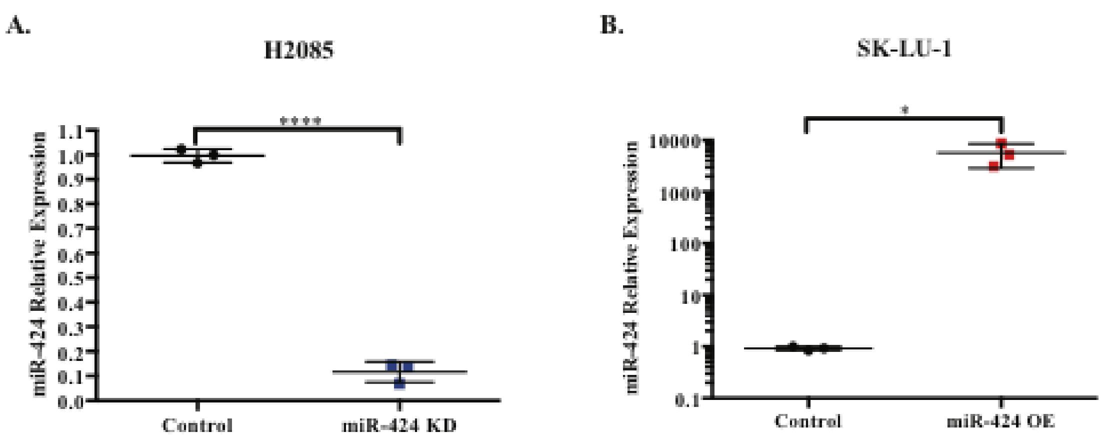

Figure 3.10: Transfection of miR-424 mimic or anti-miR-424 dramatically changes miR-424 expression levels. qRT-PCR was used to measure (**A**) miR-424-5p knockdown efficiency in H2085 lung adenocarcinoma cells ($p<0.00001$) and (**B**) miR-424 overexpression efficiency in SK-LU-1 lung adenocarcinoma cells ($p<0.05$).

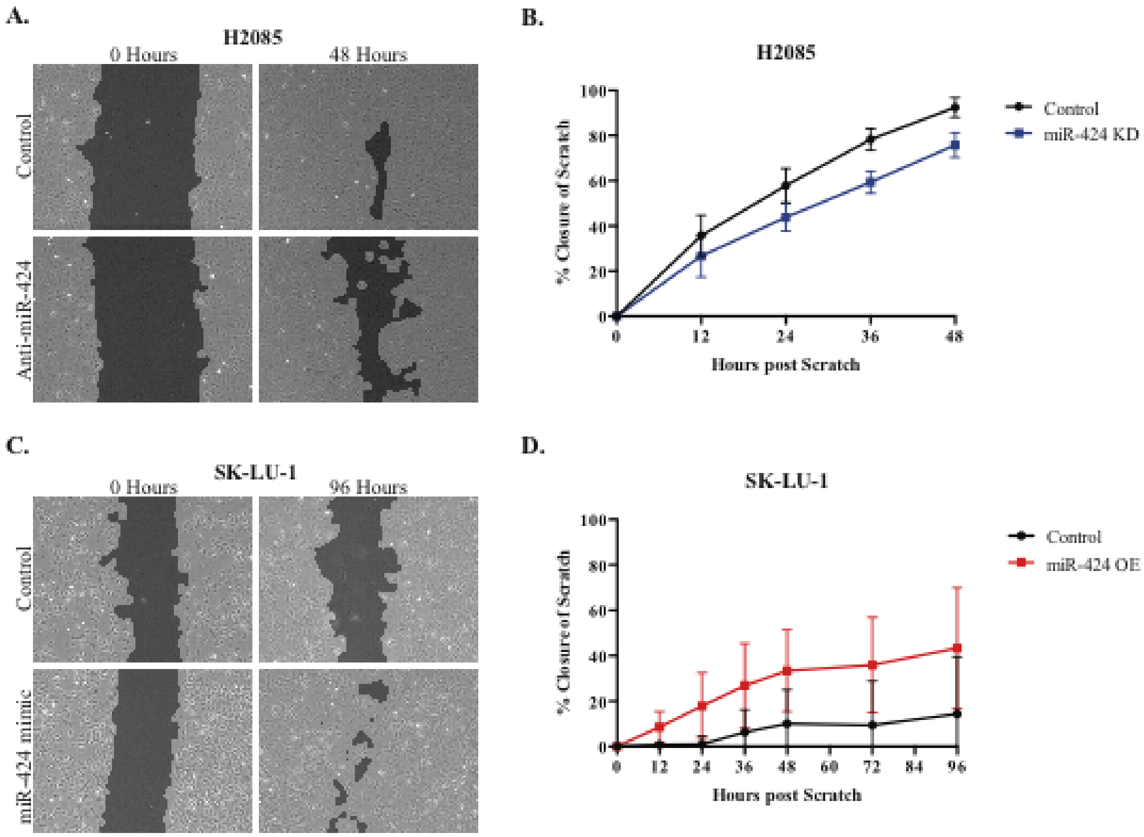

Figure 3.11: Perturbations of miR-424-5p in lung adenocarcinoma cell lines result in changes in migration. Cells are transfected with a control, anti-miR-424-5p, or miR-424 mimic. Photos are taken every 12 hours, shown are quantified results of the photos. Percent closure is calculated with TScratch. (**A**) Representative images of H2085 never smoker lung adenocarcinoma cell line shows a decrease in cell migration with miR-424 knockdown while (**B**) quantification of images over time show the change to be significant (p-value <0.05, n=3 experiments). (**C**) Representative images of SK-LU-1 smoker lung adenocarcinoma cell line shows an increase in cell migration with miR-424 overexpression and (**D**) quantification of the images show a strong increase in migration over time (p<0.005, n=3 experiments).

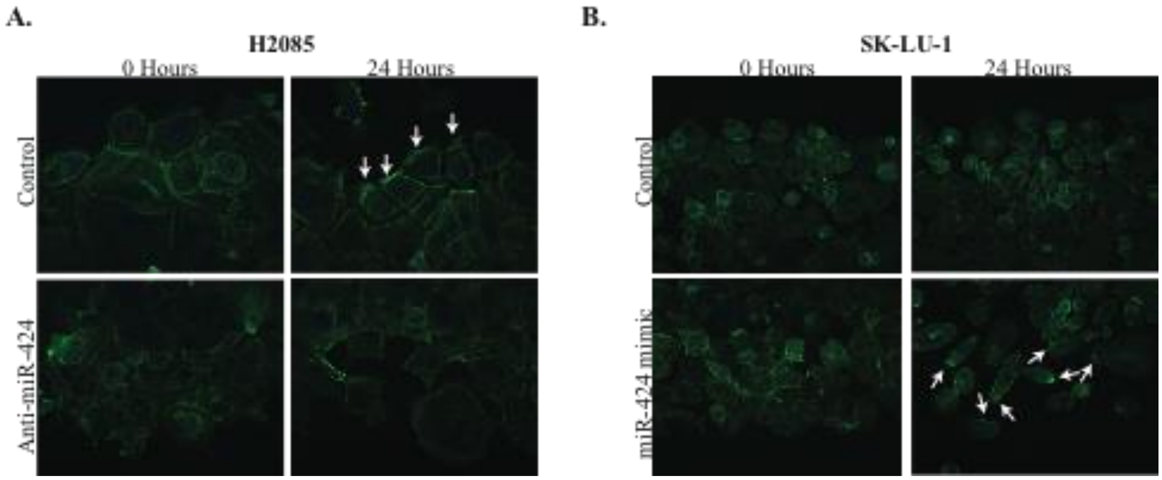

Figure 3.12: Perturbations of miR-424-5p in lung adenocarcinoma cell lines result affect the ability of the cells to form a leading edge for migration. (**A**) Staining with F-Actin shows a decrease in the number of cells with strong leading edge staining with anti-miR-424 but with (**B**) overexpression of miR-424 there is an increase in cells staining for a leading edge. Arrows point to cells with accumulated F-Actin on one side of the cell.

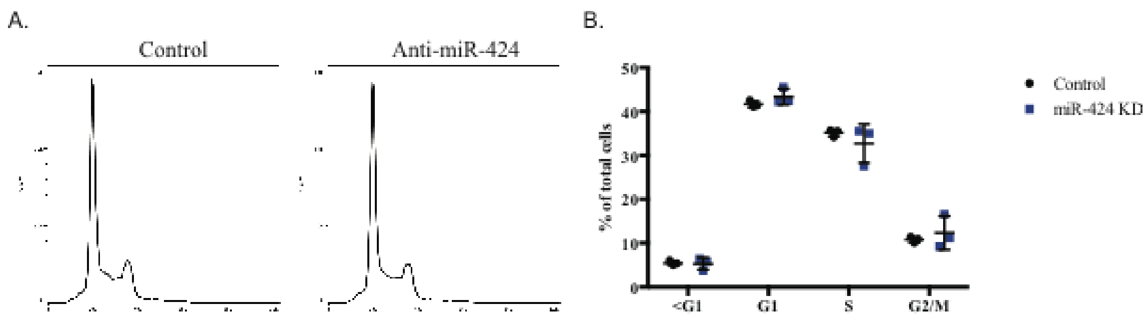

Figure 3.13: Knockdown of miR-424 does not affect cell cycle in the H2085 cell line. (**A**) Cell cycle was measured using PI staining and FACS. (**B**) Quantification of the cell cycle phases by FlowJo shows no change between control and anti-miR-424 transfected cells.

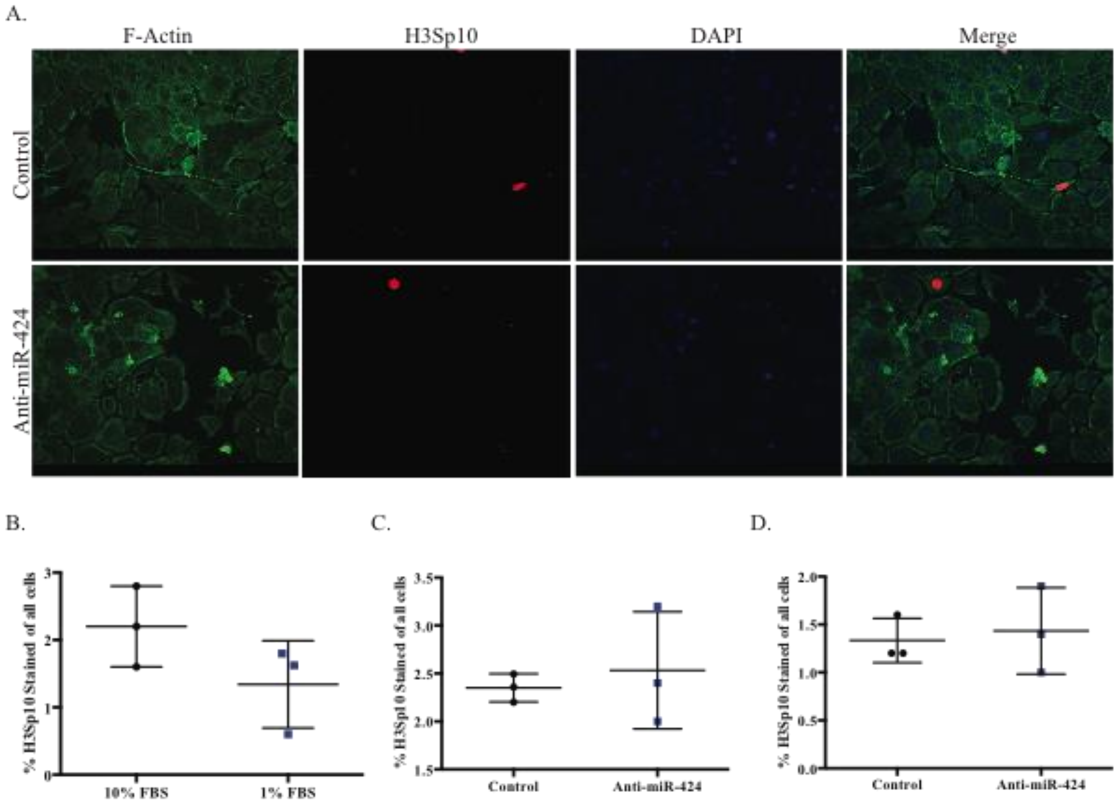

Figure 3.14: Knockdown of miR-424 does not affect mitosis in the H2085 cell line. (**A**) Staining of cells with F-Actin and H3Sp10, a marker of mitosis, show no change in the number of cells expressing the mitosis marker with around 1% of all cells expressing the marker. (**B**) Growing H2085 cells in 10% FBS (n=3) results in a slightly higher, but not significant, number of cells undergoing mitosis compared to H2085 cells grown in 1% FBS (n=3) as measured by staining for DAPI and H3Sp10. (**C**) H2085 cells grown in 10% FBS and transfected with control (n=3) or anti-miR-424 (n=3) show no difference in the number of cells in mitosis and (**D**) cells grown in 1% FBS, transfected with either control (n=3) or anti-miR-424 (n=3), and had a scratch to the cell growth surface show no difference in the number of cells in mitosis.

3.3 Materials and Methods

Tumor and Adjacent Normal Adenocarcinoma Samples

We received frozen lung adenocarcinoma tumor and adjacent normal tissues, 6 never smoker patients and 13 ever smoker patients, from the tumor biorepository at the Mayo Clinic. These samples were obtained from the resected lobes of never smokers and ever smokers with lung adenocarcinoma, where tumor cell purity was ≥70%. Normal lung tissue was obtained from an area most distant from the tumor. Additional samples were obtained from early stage (I-IIIA) resected lung adenocarcinoma collected and processed between 2011 and 2013 at the University of Texas MD Anderson Cancer Center. All 33 cases (14 never smoker patients and 19 ever smoker patients) from the MD Anderson Cancer Center included samples of lung adenocarcinoma and uninvolved normal lung parenchyma. IRB approval was obtained at all sites through an institutional ethics board and all study participants signed an informed consent form.

Small Airway Brushings

From the MD Anderson patients (14 never smoker patients and 19 ever smoker patients), adjacent (to tumor) histologically normal airway epithelial cells were obtained by brushing bronchiolar structures using sterile Cytosoft cytology brushes (Medical Packaging Corporation, Camarillo, CA) as previously described. Airway brushings were placed in Qiazol lysis buffer (Qiagen, Valencia, CA) in dry ice and immediately stored in -80°C. To confirm the content of airway epithelial cells, pan-cytokeratin immunohistochemical analysis (IHC) and analysis of neoplastic or pre-neoplastic cells were performed as previously described.

qRT-PCR

To measure the expression of miR-424-5p, 10 ng of total RNA was used in a Taqman MiRNA Assay (Life Technologies, Catalog #4427975, ID #000604, Carlsbad, CA) as per manufacturer's protocol and the results were normalized to RNU44 expression (Life Technologies, Catalog #4427975, ID #001094, Carlsbad, CA). To measure the expression of *FOXP2*, 500 ng of total RNA was reverse transcribed using RT2 First Strand Kits (Qiagen, Catalog #330401, Valencia, CA) according to the manufacturer’s protocol. cDNA product was added to SYBR Green qPCR Mastermix (Qiagen, Catalog #330523, Valencia, CA) and *FOXP2* primer (Qiagen, Catalog #PPH01974B, Valencia, CA). Data was normalized to the expression of UBC (Qiagen, Catalog #PPH00223F, Valencia, CA) and analyzed using the comparative CT method. Relative expression was compared using a student’s t-test. For both comparing qRT-PCR fold change of tumor vs adjacent normal for FOXP2 and miR-424-5p between ever and never smokers as well as for comparing fold change between EGFR status types, we utilized a Wilcoxon test.

Cells and Cell Culture

Never smoker lung adenocarcinoma cell line H2085 (obtained from ATCC, verified by ATCC, Manassas, VA) were grown in DMEM:F12 media supplemented with 0.02 mg/ml insulin, 0.01 mg/ml transferrin, 25 nM sodium selenite, 50 nM hydrocortisone, 1 ng/ml epidermal growth factor, 0.01 mM ethanolamine, 0.01 mM phosphorylethanolamine, 100 pM triiodothyronine, 0.5% BSA, 10 mM HEPES, 0.5 mM sodium pyruvate, 2 mM L-glutamine and 10% FBS. Cells were subcultured as described by ATCC. The lung adenocarcinoma cell line SK-LU-1 (obtained from ATCC, verified by ATCC, Manassas, VA) was grown in EMEM supplemented with 10% FBS and subcultured as described by ATCC.

Migration Assay

H2085 was transiently transfected with hsa-miR-424-5p miRCURY LNA miRNA inhibitor or miRCURY LNA miRNA inhibitor control (Exiqon, Catalog #4104094-001, 199006-001, Woburn, MA). SK-LU-1 was transiently transfected with hsa-miR-424-5p *mir*Vana miRNA mimic (Life Technologies, Catalog #4464066, ID #MC10306, Carlsbad, CA) or Negative Control #1 *mir*Vana miRNA mimic (Life Technologies, Catalog #4464058, Carlsbad, CA). Cells were grown to full confluence and then serum-starved in RPMI-1640 supplemented with 1% FBS for 24 hours. The plate surface was then scratched with a 200 µl tip and cells were kept in a serum-starved environment for the duration of the assay. Photographs of the scratch were taken every 12 hours at 50x magnification using the Zeiss inverted microscope, Axio Vert.A1 and the microscopy camera Axiocam 105 color. Photographs were taken using AxioVision SE64 Rel. 4.9.1 SP1 software. Tiff images were analyzed using Matlab Compiler Runtime v7.8 with TScratch application v1.0.

Cell Cycle Assay

H2085 cells were grown and transiently transfected as previously described. Cells were collected 96 hours after transfection and fixed in 66% ethanol on ice. At the time of analysis cells were stained with 1X PI and RNase staining solution and incubated in the dark for 30 minutes. Cells were then examined with a FACS flow cytometer, FACSCalibur (BD Biosciences, San Jose, CA) and DNA histograms were analyzed with FlowJo v10.0.8. Each experimental condition was repeated in triplicate.

Cell Staining

Cells were fixed with 4% paraformaldehyde and permeabilized with 0.2% Triton X. Cells stained for anti-histone H3 (phospho S10) (abcam, ab14955, Cambridge, MA) were blocked with 1% BSA and incubated overnight with the primary antibody. Cells were incubated with Alexa Fluor 546 goat anti mouse for two hours at room temperature

(Life Technologies, A11003, Carlsbad, CA). Alexa Fluor 488 Phalloidin (Life Technologies, A12379, Carlsbad, CA) was used to stain for F-Actin and ProLong Gold Antifade Mountant with DAPI (Life Technologies, P-36931, Carlsbad, CA) was used to mount and stain the cells for DAPI. Fluorescence was imaged at 40x using a Zeiss AxioSkop 40 microscope and a Zeiss AxioCam MRm black and white camera. Images were collected using the Axio Vision software (Release 4.6.3 SP1, 11-2007) and analyzed using ImageJ 1.43u software.

Microarrays

Total RNA from a knockdown study in H2085 cells (3 control, 3 knockdown) was isolated using the miRNeasy kit (Qiagen, 217004, Valencia, CA) and profiled by microarray (using the Affymetrix GeneChip Human Transcriptome Array HTA 2.0 platform). CEL files were normalized to produce gene-level expression values using the implementation of the Robust Multi-array Average (RMA) in the *affy* package (version 1.36.1) included in the Bioconductor software suite, and an Entrez Gene-specific probeset mapping (v16.0.0) from the Molecular and Behavioral Neuroscience Institute (Brainarray) at the University of Michigan. RMA was performed using the R environment for statistical computing (v2.15.1). Array quality was assessed by computing Relative Log Expression (RLE) and Normalized Unscaled Standard Error (NUSE) metrics using R and the *affyPLM* Bioconductor package (v1.34.0). RLE, NUSE, and PCA were used for outlier detection. Gene expression differences between knockdown and control were assessed using a t-test. Gene Set Enrichment Analysis (GSEA) was used to test for enrichment of a ranked gene list ranked based on the t-statistic of control versus experimental condition against selected gene sets from MSigDB.

3.4 Discussion

Building upon previous work showing transcriptomic differences between ever and never-smoker lung adenocarcinoma, we have identified one key driver microRNA, miR-424, which impacts cancer cell migration *in vitro*.

MicroRNAs play a key role in regulating pathways and promoting oncogenic activity [135,136]; they also serve as attractive molecular targets for cancer therapies. For this reason, by finding differentially expressed microRNAs involved in cancer, we are identifying microRNAs that can be further pursued as potential therapeutic targets. Analyzing the previously produced never smoker adenocarcinoma regulatory network, we were able to gain new insights into the role that microRNAs may play at the regulatory level to promote carcinogenesis in never smokers. It is through this analysis that miR-424 warranted our attention as a key hub in the network, exercising potential regulatory interactions with a large number of mRNA targets that participate in many pathways, particularly those deregulated in cancer, such as migration. Furthermore, miR-424 itself has been reported to play an oncogenic role in colonic and pancreatic adenocarcinoma [98,99]. Importantly, the unique differential expression of miR-424 in never smoker tumor versus adjacent normal was validated experimentally using qRT-PCR. The changes in miR-424 expression are independent of sample *EGFR* mutational status, implying that miR-424 works in different pathways then *EGFR* and its mechanism of action in cancer may expand our understanding of targetable pathways for therapeutic intervention.

Building on the success of identifying a potential regulatory microRNA, we knocked down miR-424 in a never smoker lung adenocarcinoma cell line and detected perturbations in multiple migration pathways using gene expression arrays. To validate these findings, we proved the critical importance of miR-424 as a driver of a key cancer hallmark by demonstrating that knockdown of miR-424 in a never smoker cancer cell line (H2085) robustly and reproducibly reduces cancer cell migration. Additionally, in an adenocarcinoma cell line with little endogenous migratory potential (SK-LU-1) we overexpressed miR-424 and found the microRNA is capable of promoting cell migration. The miR-424 ability to modulate migration in vitro was supported by several lines of evidence, F-Actin staining demonstrated that the organized leading edge of H2085 cells upon the introduction of a scratch is lost with miR-424 knockdown. Conversely, SK-LU-1 cells don't present with leading edges until they are transfected with miR-424 mimic. Collectively, these experiments argue a role for miR-424 in cellular migration in lung adenocarcinoma. It must be noted, however, that many of the lung adenocarcinoma cell lines tested for a migration phenotype were not affected by knockdown of miR-424. Expression of miR-424 may have been low in some of these cell lines to begin with, preventing knockdown of miR-424 from having a strong reaction, or they may not be dependent on miR-424 for migration. SK-LU-1 is an example where the knockdown of miR-424 did not alter migration but overexpression did. Additional cell lines need to be tested for both miR-424 knockdown and miR-424 overexpression in regards to migration to better understand the specific adenocarcinoma tumors that can be manipulated with miR-424 perturbations. Beyond the scope of this current work, it would be compelling to test if miR-424 plays a role in hallmarks of cancer other than migration, using both additional cell line assays and mouse models, as well as identify the specific pathways modulated with miR-424 perturbations that are responsible for the migration phenotype so that we can identify a targetable pathway in lung adenocarcinoma.

Others have described local transcriptomic changes in tissue and the epithelium of small airways located proximal to the tumor in smokers [137] as well as in the epithelium of large airways of smokers [138]. These seminal observations lead to the development of a minimally invasive bronchoscopy based diagnostic test for lung cancer in current or former smokers[139]. Very little is known about the field effect in never smokers with lung cancer. One study found *EGFR* mutations in the histologically normal bronchial and bronchiolar epithelia of patients with lung adenocarcinoma bearing *EGFR* mutations[140]. This study includes never smoker patients as well as former smokers, with no never smoker specific results highlighted. More recently, an article of the adjacent airway field of cancerization in NSCLC noted that, while they did not explore the field of never smoker lung adenocarcinoma, this area is relatively in need of future research[137]. In our study we present preliminary observational data that suggests that the high miR-424 expression in never smoker tumors extends into the small airways of an independent cohort of never smokers with lung adenocarcinoma, raising the possibility of a transcriptomic field effect specific to never smokers. If miR-424 is indeed expressed more highly in the airways of never smokers with lung cancer as compared to their lung cancer free counterparts it suggests the possibility of being able to develop a minimally invasive diagnostic test for lung cancer in never smokers. Our observation of a higher expression of miR-424 in the never smoker adjacent to tumor small airway has important implications for developing diagnostic tests for this cohort, but needs to be confirmed by profiling miR-424 in the small airways of never smokers without lung cancer as well as moving from the small airway, close to the tumor, to the large airways.

3.5 Conclusions

In conclusion, our results suggest that oncomiR miR-424 may be a therapeutic target in never smoker lung adenocarcinoma. Migration, a hallmark of cancer, is affected by miR-424 perturbations. By further exploring the mechanism by which miR-424 affects migration, we can potentially identify therapeutics to counteract the role of miR-424 in this disease. Although our results need to be confirmed with controls, the observation of miR-424 expression in the small airways near never smoker lung adenocarcinoma tumors combined by the already proven feasibility of transcriptome based biomarker development suggests yet another rapid path to clinical impact for our results. Validation of these profiles, including the activation of miR-424 as a potential oncomiR in the tumor versus adjacent normal of adenocarcinoma patients with no history of smoking, represents critical findings that may ultimately contribute to the clinical development of new therapies and diagnostics for lung adenocarcinoma in never smokers.

CHAPTER FOUR

MicroRNA 34c 5' IsomiR Regulates Ras Signaling Genes in Interstitial Lung Disease

4.1 Introduction

Patients with interstitial lung diseases are characterized as having thickening of the supporting tissues between the alveoli of the lung[141]. Thickening can be due to a number of factors such as autoimmune diseases, exposure to toxins such as tobacco smoke, genetic diseases, or an unknown cause (idiopathic)[46,142]. A definitive diagnosis of a patient's ILD is critical to identify a treatment, since they vary from antimicrobial agents, chemotherapy, avoidance behavior, to only palliative care[44,141]. The most common form of ILD is IPF, which has a median survival of three years from time of diagnosis and has no effective therapy available[48,143–145].

Currently, ILD is described at the tissue level with processes such as apoptosis, fibrosis, aberrant wound repair, and scaring[146,147]. At the molecular level ILD is associated with dysregulation of TGF-ß and Wnt signaling pathways, extracellular matrix remodeling, and epithelial to mesenchymal cell transition[148–150]. Linkage studies have indicated dysregulation of MUC5B is associated with IPF[151].

Very little is known about the role of microRNAs in ILD, with most of the focus on microRNAs that are associated with TGF-ß[148–150]. Because microRNAs perform critical functions as gene regulators in healthy biology as well as in many disease processes, it is important to characterize the microRNA role of ILD in relation to the disease onset and progression.

As part of the LGRC, our group previously sequenced the small RNA in lung tissue samples from patients with COPD (n = 150) or ILD (n = 149) and from normal lung (n=65) (Table 4.1).

	Control (n=65)	ILD (n=149)	COPD (n=150)
Smoking Status†‡	2 Current, 38 Former, 20 Never, (5)	5 Current, 89 Former, 51 Never, (4)	8 Current, 134 Former, 6 Never, (2)
Age†‡	63.0 +/- 12.0	61.2 +/- 10.2	64.3 +/- 10.1
Pack Years*†‡	41.1 +/- 36.6 (24)	26.0 +/- 19.6 (55)	55.2 +/- 38.5 (8)
Gender	32 Male, 33 Female	80 Male, 69 Female	88 Male, 62 Female
FEV_1/FVC*†‡	0.78 +/- 0.1	0.83 +/- 0.1 (2)	0.50 +/- 0.2 (1)
Percent Emphysema†‡	0.66 +/- 1.0 (26)	0.77 +/- 1.7 (60)	16.4 +/- 17.8 (30)

Table 4.1: Demographics of LGRC cohort sequenced with microRNA sequencing. The symbol * denotes significant difference ($p<0.05$) between the control and ILD cohorts. † denotes significant difference ($p<0.05$) between the control and COPD cohorts. ‡ denotes significant difference ($p<0.05$) between the ILD and COPD cohorts. Numbers in brackets denote samples with missing information in that field.

Our group found that the expression of 255 microRNAs were significantly associated with the presence of disease (FDR < 0.05; absolute fold change > 1.25; Figure 4.1A) while controlling for smoking status, age, total number of reads aligned to microRNA loci within a sample, and library preparation protocol. To characterize the heterogeneity observed within these disease-associated microRNAs, our group used consensus clustering to uncover five distinct subgroups of samples (s1-s5). Cluster s1 contained the majority of control samples (33 out of 62; Figure 4.1B). Cluster s3 contained the largest proportion of COPD patients while clusters s4 and s5 contained the largest proportions of ILD patients. Several clinical measures of disease severity were associated with sample cluster status with each disease including DLCO in ILD and DLCO, FEV1 percent predicted, FEV1/FVC, percent emphysema, and BODE in COPD ($p < 0.05$; Figure 4.1C).

Looking closely at the microRNA clusters identified by consensus clustering, m1-m4 (Figure 4.1A), microRNA cluster m1 contained upregulated microRNAs in IPF clusters s4 and s5. The microRNAs in m1 are enriched for microRNAs located in the cluster on 14q32 as well as other microRNAs previously implicated in IPF such as miR-21 and miR-154[73,152]. The microRNAs in cluster m2 are primarily responsible for the separation observed in ILD-associated clusters s4 and s5. Many of the m2 microRNAs are specifically expressed in airway epithelium and involved in airway differentiation including miR-449a/b/c, miR-34b/c, miR-4423, and miR-205[94]. MicroRNAs in m3 are primarily up-regulated in the COPD-enriched cluster s3 and included miR-15/16, miR-144/451, as well as some microRNAs in the miR-17-92 cluster. Finally, microRNAs in cluster m4 included miR-30a/c/d and miR-218.

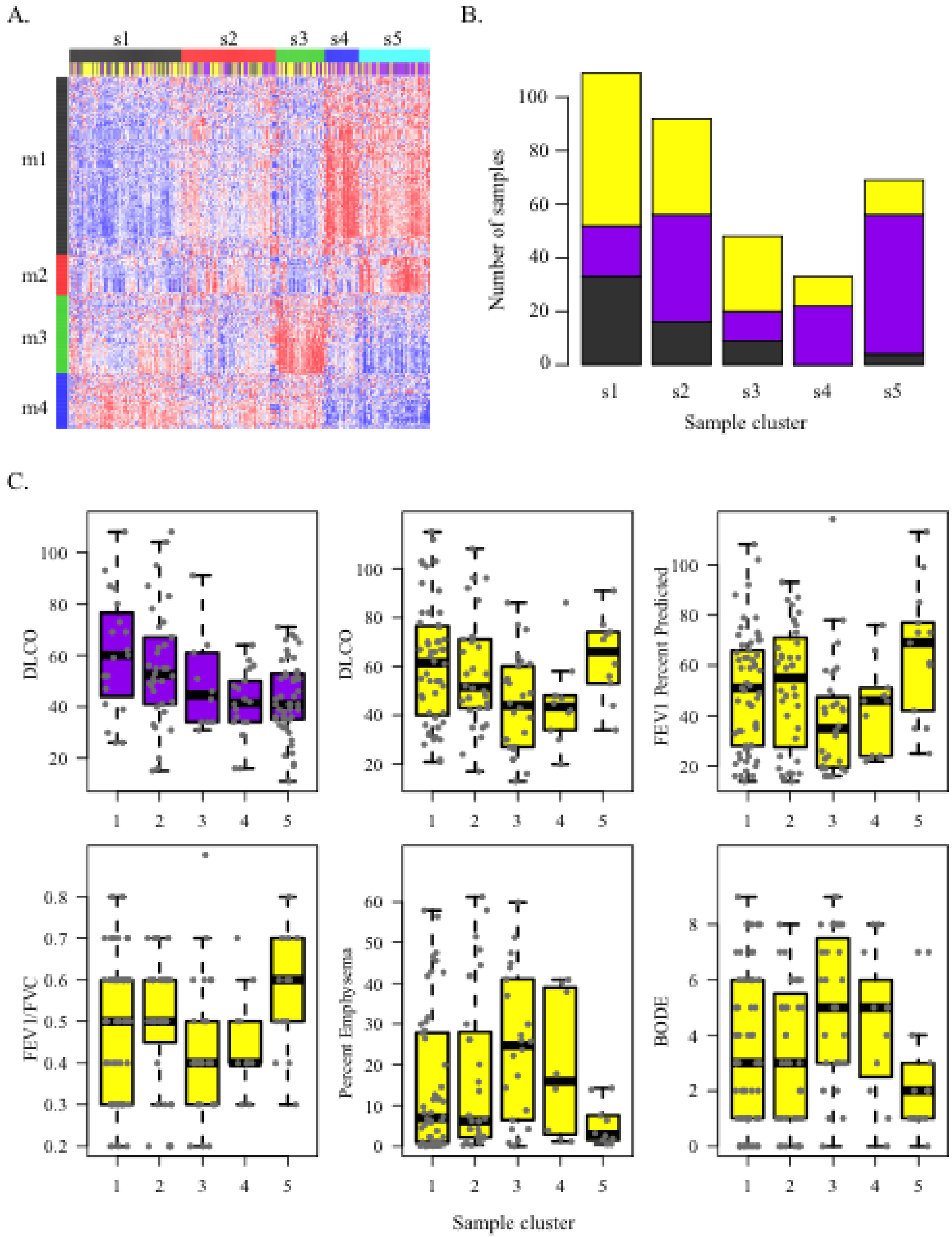

Figure 4.1: MicroRNA sequencing analysis of COPD and ILD tissue shows distinct clusters of microRNAs associated with the different diseases compared to control. (**A**) A total of 255 microRNA were identified as significantly associated with the presence of disease and using consensus clustering five disease subgroups, denoted in the first color bar with s1-s5, and four microRNA clusters, denoted on the left side of the heat map with m1-m4, were discovered. The second color bar shows the patient population, with black bars for control, yellow bars for COPD, and purple bars for ILD patients. (**B**) The five patient subgroups vary in their distribution of number of patients with each disease or control phenotype. (**C**) Many clinical measurements are associated with the subgrouping, including DLCO in ILD as well as DLCO, FEV1 percent predicted, FEV1/FVC, percent emphysema, and BODE in COPD ($p<0.05$).

IsomiRs were recently identified and described as variants of what is known as the canonical microRNA sequence. These variants highlight that what is commonly known as the canonical form of a microRNA, identified by being the most highly expressed form of the microRNA via sequencing, is sometimes not the highest expressed form of the microRNA in all tissue and disease types[153,154]. Most isomiRs are variants on the 3' end of the microRNA. A variant on the 5' end of a microRNA, shifts the seed sequence, potentially affecting the targets of the microRNA[154,155].

To evaluate the potential role of microRNA isoforms with alternative seed sequences, reads were grouped according to their aligned start position within the microRNA precursor region and those with the different start positions were counted as separate transcripts. Our group focused on microRNAs in cluster m2 due to their ability to distinguish sample clusters s4 and s5 as well as their known roles in airway differentiation. The miR-449/34 family had some of the highest expressed non-canonical isomiRs in this cluster. miR-34c was upregulated in ILD patients and had a highly expressed 5' isomiR (Figure 4.2).

miR-34c canonical	AGGCAGTGTAGTTAGCTGATTGC
miR-34c 3' isomiRs	AGGCAGTGTAGTTAGCTGATT
	CGGCAGTGTAGTTAGCTGATTGC
	AGGCAGTGTAGTTAGCTGATTGCT
	AGGCAGTGTAGTTAGCTGATTG
	AGGCAGTGTAGTTAGCTGATTGT
	AGGCAGTGTAGTTAGCTGATTGA
	AGGCAGTGTAGTTAGCTGAT
	AGGCAGTGTAGTTAGCTGA
	AGGCAGTGTAGTTAGCTGATTGCA
miR-34c 5' isomiR	TAGGCAGTGTAGTTAGCTGATT

Figure 4.2: The top miR-34c variants by read count with RNA sequencing. The seed sequence (red) does not change with 3' isomiRs but is shifted with the 5' isomiR.

Using TargetScan, predicted targets of miR-34c and the miR-34c 5’ isomiR were identified. Our group also performed all pairwise Spearman correlations between the expression of every gene and every microRNA. The distribution of genes predicted to have target sites for miR-34c canonical seed, the miR-34c 5’ isomiR seed, or both are enriched for genes that are anticorrelated to the microRNAs, compared to a control distribution of genes that are not predicted to be targets ($p<0.01$; Wilcoxon rank-sum test; Figure 4.3A). In order to further classify the biological processes targeted by miR-34c and the miR-34c 5’ isomiR, their individual as well as combined predicted targets were examined for pathway enrichment. Ras protein signaling was one of the top pathways enriched among genes that were significantly anti-correlated to miR-34c-5p locus expression ($FDR<0.25$) and that had a predicted binding site for the isomiR only ($p<0.0005$; Figure 4.3B). In addition, Ras protein signaling was also modestly enriched among anti-correlated genes that were predicted to be targets of both miR-34c and isomiR ($p<0.05$; Figure 4.5B). However this pathway was not enriched in anti-correlated miR-34c canonical-only targets.

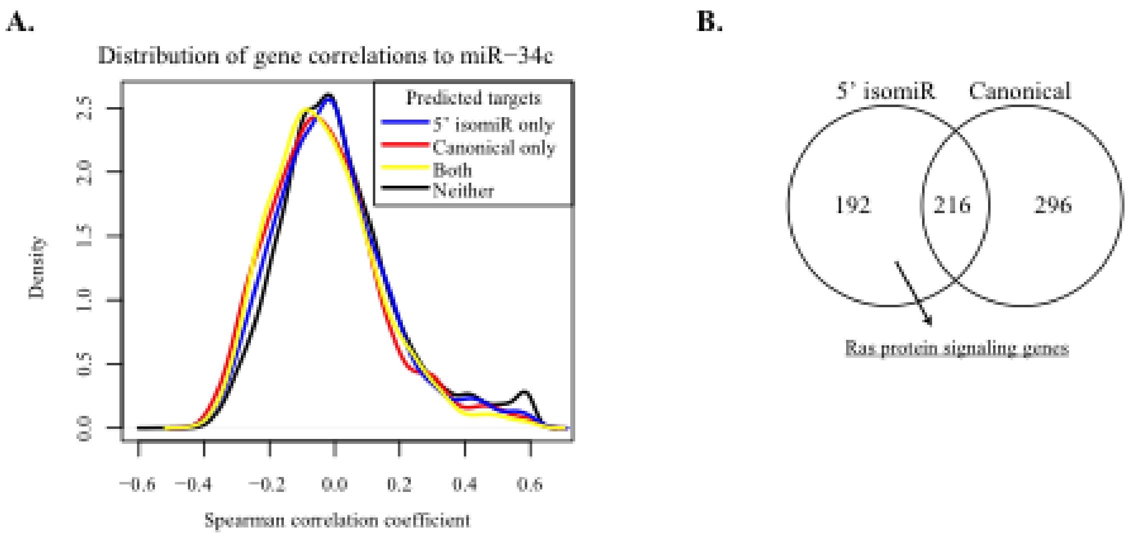

Figure 4.3: Predicted targets of miR-34c and miR-34c 5' isomiR include individual and co-targets that regulate Ras protein signaling. (**A**) The distribution of Spearman correlation coefficients between miR-34c-5p expression and gene expression is plotted for four different sets of genes: predicted targets of the 5' miR-34c isomiR (blue), the canonical miR-34c (red), both (yellow), and neither (black). (**B**) Significantly anti-correlated (FDR<0.25) predicted targets of the 5' miR-34c isomiR are enriched for Ras protein signaling genes (p<0.0005).

Based on the data generated in these previous studies, we aim to characterize miR-34c, identified by RNA sequencing to be upregulated in ILD tissue compared to healthy tissue. miR-34c is not only dysregulated in ILD, but by sequencing we see that there is a 5' isomiR of miR-34c that expands the potential targeting breadth for miR-34c. By modulating the expression of miR-34c and its 5' isomiR, we seek to identify common and independent mRNA targets for these two variants.

4.2 Results

4.2.1 Validation of miR-34 expression in clinical samples

The behavior of miR-34c as well as the behavior of the other two miR-34 family members (miR-34a and miR-34b) was validated by qRT-PCR in a subset of patient samples (10 control, 10 ILD; Table 4.2). The differential expression of miR-34c between control and ILD patient samples was confirmed by qRT-PCR (Figure 4.4A) but there was no significant difference in the expression of the other family members, miR-34a (Figure 4.4B) and miR-34b (Figure 4.4C).

	Control (n=10)	**ILD (n=10)**
Smoking Status	1 Current, 5 Former, 4 Never	0 Current, 8 Former, 2 Never
Age	56.1 +/- 15.2	63.1 +/- 4.4
Pack Years	29.4 +/- 20.9 (3)	26.9 +/- 26.0 (2)
Gender	5 Male, 5 Female	7 Male, 3 Female
FEV_1/FVC	0.78 +/- 0.08	0.8 +/- 0.08
Percent Emphysema	0.23 +/- 0.21 (5)	2.06 +/- 2.64 (5)

Table 4.2: Demographics of LGRC samples used in qRT-PCR studies.

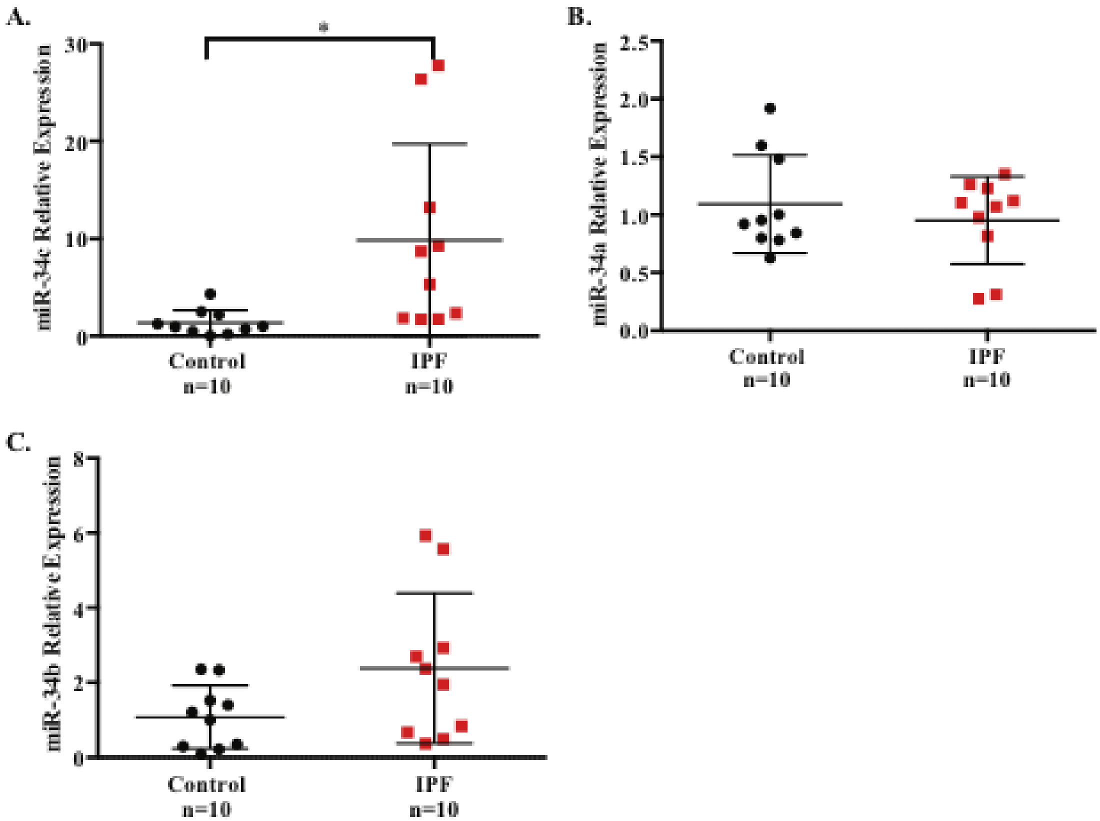

Figure 4.4: Upregulation of miR-34c in IPF tissue validates by qRT-PCR but miR-34a and miR-34b are not differentially expressed. (**A**) IPF (n=10) tissues have a significantly higher expression of miR-34c than control (n=10) tissues (p<0.05). The expression of (**B**) miR-34a is unchanging between IPF (n=10) and control (n=10) tissue while (**C**) miR-34b appears to have a small, but not significant, increase in expression in IPF tissue compared to control.

4.2.2 Validation of miR-34c and miR-34c 5' isomiR predicted targets in clinical samples

In order to validate the Ras protein signaling pathway genes identified as being predicted targets of miR-34c and/or the miR-34c 5' isomiR, the expression of some of these genes was measured in clinical samples via a secondary platform, qRT-PCR. *NOTCH1*, a gene that has previously been described as a direct target of miR-34c and is a predicted target of the miR-34c 5' isomiR is not a member of the RAS protein signaling pathway but serves as a positive control. *NOTCH1* (p<0.0005; Figure 4.5J), as well as *GRB2* (p<0.00005; Figure 4.5B), *CRK* (p<0.005; Figure 4.5C), *CRKL* (p<0.0005; Figure

4.5H), *RHOC* ($p<0.005$; Figure 4.5F), *RALA* ($p<0.05$; Figure 4.5A) and *GRAP* ($p<0.0005$; Figure 4.5D) all validate with significant downregulation of their expression in ILD samples compared to the control samples. Genes *RHOA* (Figure 4.5E), *ARAP2* (Figure 4.5G), and *EGF* (Figure 4.5I) do not validate as significantly differentially expressed.

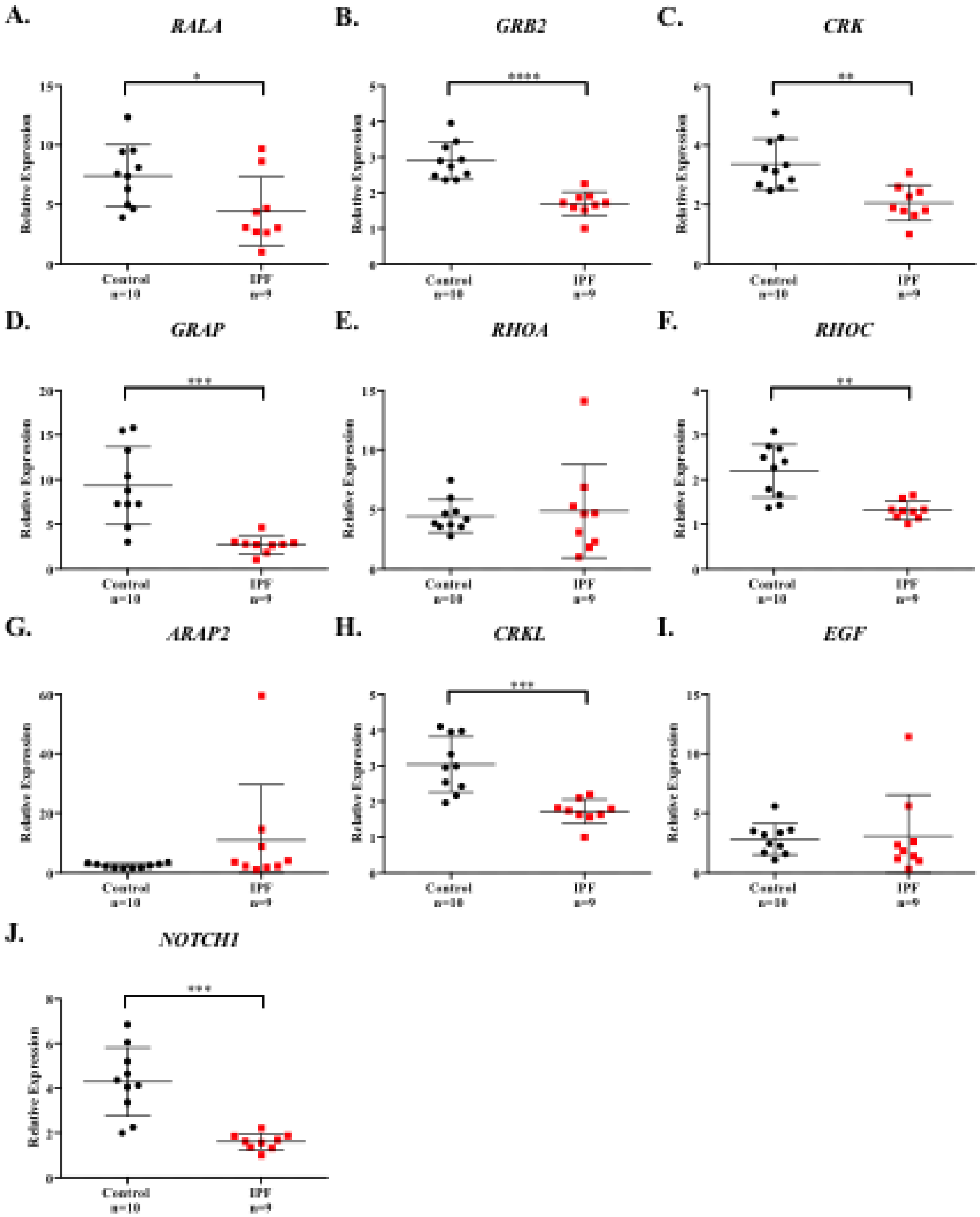

Figure 4.5: Predicted targets of miR-34c variants validate as being dowregulated in ILD clinical tissue. Of the ten genes measured in clinical samples, nine of which are members of the Ras protein signaling pathway, seven validated as being downregulated in ILD compared to control tissue. These seven genes are: (**A**) *RALA* ($p<0.05$), (**B**) *GRB2* ($p<0.00005$), (**C**) *CRK* ($p<0.005$), (**D**) *GRAP* ($p<0.0005$), (**F**) *RHOC* ($p<0.005$), (**H**) *CRKL* ($p<0.0005$), and (**J**) *NOTCH1* ($p<0.0005$).

4.2.3 Lung fibroblast cell line IMR90 allows for miR-34c regulation of NOTCH1

Lung fibroblast (IMR90) and basal airway (HBEpC) cells were transfected with mimics of miR-34c, the miR-34c 5' isomiR, or both in order to identify targets of these microRNA. Leveraging the knowledge that *NOTCH1* has been well described as a direct target of miR-34c[156,157] and in our data is a predicted target of the miR-34c 5' isomiR, it was the first gene profiled by qRT-PCR to identify which cell line would be best for subsequent experiments. In fibroblast cells transfection with all variations of the mimics downregulated the overall gene expression of *NOTCH1* significantly ($p<0.005$, $p<0.005$, $p<0.005$; Figure 4.6A). In basal airway cells, none of the conditions resulted in a change in *NOTCH1* expression (Figure 4.6B). Based on the results of the positive control, all additional experiments were done in IMR90 cells.

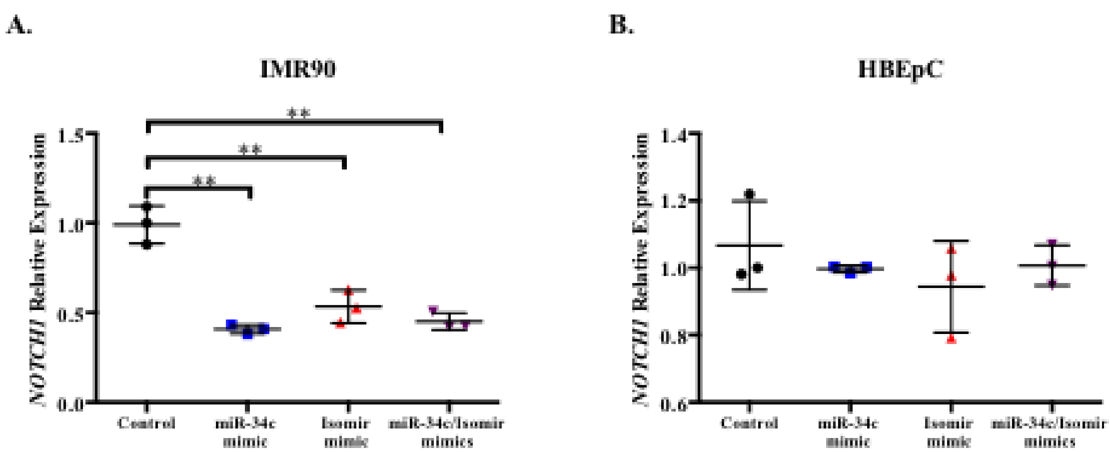

Figure 4.6: IMR90 fibroblast cells are sensitive to miR-34c transfections as evidenced by positive control *NOTCH1* gene expression. Cells transfected with miR-34c mimic, miR-34c 5' isomiR mimic, or both were measured for *NOTCH1* gene expression. (**A**) IMR90 fibroblast cells show significant repression of *NOTCH1* with all experimental transfections ($p<0.005$, $p<0.005$, $p<0.005$). (**B**) HBEpCs show no significant change in *NOTCH1* expression compared to control.

4.2.4 MiR-34c 5' isomiR has targets independent of the canonical miR-34c

In IMR90 cells transfected with mimics of miR-34c, the miR-34c 5' isomiR, or both miR-34c and the miR-34c 5' isomiR, there are members of the Ras protein signaling pathway that are affected only by transfection with the miR-34c 5' isomiR mimic. Of the genes that validate as being downregulated with miR-34c 5' isomiR mimic transfection, *GRB2* has the greatest gene expression downregulation ($p<0.005$; Figure 4.7A) and *CRKL* has a smaller change ($p<0.05$; Figure 4.7B). To validate that we are in fact seeing a change in gene expression with miR-34c 5' isomiR overexpression and not an artifact, *NOTCH4* was measured in the same samples. *NOTCH4* is a known target of miR-34c[158] and is not predicted to be targeted by the miR-34c 5' isomiR. Expression of *NOTCH4* is significantly decreased with miR-34c overexpression ($p<0.005$; Figure 4.7C).

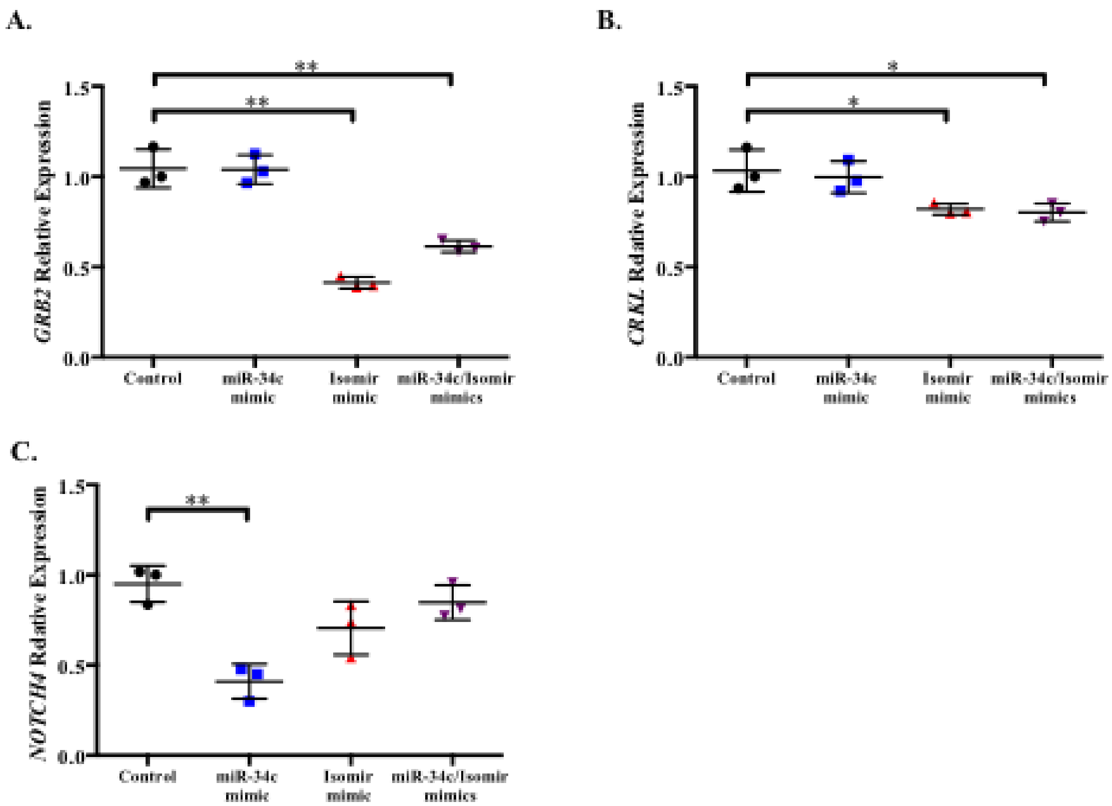

Figure 4.7: The miR-34c 5' isomiR and the canonical miR-34c have independent targets. (**A**) *GRB2* is significantly downregulated with the miR-34c 5' isomiR mimic transfection ($p<0.005$, $p<0.005$) but not the miR-34c canonical transfection. (**B**) *CRKL* is also significantly downregulated with the miR-34c 5' isomiR mimic transfection ($p<0.05$, $p<0.05$) but not the miR-34c canonical transfection. (**C**) *NOTCH4* expression is decreased with miR-34c mimic transfection ($p<0.005$).

4.2.5 Expression of miR-34 family members are affected by overexpression of miR-34c

To determine if other members of the miR-34 family were affected by flooding the cells with miR-34c and 5' miR-34c isomiR mimics, miR-34a and miR-34b expression were measured by qRT-PCR. The expression of miR-34b measures as significantly increased with miR-34c mimic transfection, this is most likely an artifact of the primer measuring miR-34c expression as well as miR-34b ($p<0.000005$, $p<0.0005$; Figure 4.8A). Interestingly, miR-34a expression is reduced with transfection of any of the mimics ($p<0.0005$, $p<0.005$, $p<0.005$; Figure 4.8B).

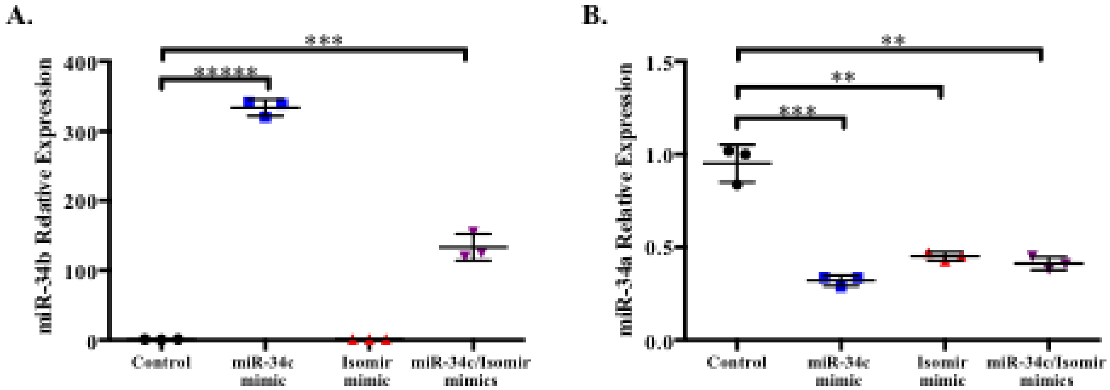

Figure 4.8: MiR-34 family members are affected by miR-34c and miR-34c 5' isomiR overexpression. (**A**) The expression of miR-34b is increased with miR-34c overexpression ($p<0.000005$, $p<0.0005$). (**B**) miR-34a expression is downregulated with both miR-34c and the miR-34c 5' isomiR transfections ($p<0.0005$, $p<0.005$, $p<0.005$).

4.3 Materials and Methods

Control and ILD Samples

A subset of samples obtained from the NHLBI funded LTRC[159] that we performed small RNA sequencing on were used to perform qRT-PCR validation.

qRT-PCR

To measure the expression of miR-34a-5p, miR-34b-5p and miR-34c-5p, 10 ng of total RNA was used in a Taqman MiRNA Assay (Life Technologies, Catalog #4427975, ID #000426, 000427, 000428, Carlsbad, CA) as per manufacturer's protocol and the results were normalized to RNU44 expression (Life Technologies, Catalog #4427975, ID #001094, Carlsbad, CA). To measure the expression of *RALA*, *GRB2*, *CRK*, *CRKL*, *GRAP*, *RHOA*, *RHOC*, *EGF*, *ARAP2*, *NOTCH4* and *NOTCH1*, 500 ng of total RNA was reverse transcribed using RT2 First Strand Kits (Qiagen, Catalog #330401, Valencia, CA) according to the manufacturer's protocol. cDNA product was added to SYBR Green qPCR Mastermix (Qiagen, Catalog #330523, Valencia, CA) and the appropriate primer

(Qiagen, Catalog #PPH07458A, PPH00714C, PPH00731A, PPH01982A, PPH13173A, PPH00305G, PPH01089E, PPH00137B, PPH20012A, PPH06021F, PPH00526C, Valencia, CA). Data was normalized to the expression of UBC (Qiagen, Catalog #PPH00223F, Valencia, CA) and analyzed using the comparative CT method.

Transfection

IMR90 cells and HBEpCs were transiently transfected with hsa-miR-34c-5p miRIDIAN miRNA mimic (Dharmacon, Catalog #C-300655-03-0020, Lafayette, CO), a custom miR-34c 5' isomiR miRIDIAN miRNA mimic (Dharmacon, Lafayette, CO) or miRIDIAN miRNA mimic Negative Control #1 (Dharmacon, Catalog #C-001000-01, Lafayette, CO). IMR90 cell transfection was completed using Lipofectamine RNAiMAX transfection reagent (Life Technologies, Catalog #13778150, Carlsbad, CA) according to the manufacturer's protocol. Transfection of HBEpCs was done using Cytofect Epithelial Cell Transfection Kit (Cell Applications, Catalog #TF102K, San Diego, CA).

4.4 Discussion

The sequencing of tissue from control and ILD samples led to the discovery of miR-34c being upregulated in ILD. Expanding on these studies, we have shown that not only is miR-34c upregulated in ILD patient tissues compared to controls, but that the expression of miR-34c 5' isomiR leads to regulation of members of the Ras signaling pathway.

Very little is known about the molecular underpinnings of ILD. In order to better understand and treat the diseases listed under the umbrella of ILD, the microRNAs dysregulated with the onset and progression of the disease as well as the networks of mRNA affected by these microRNAs need to be investigated. By doing this, we can better diagnose and treat ILD. The work done in this study validates the expression of miR-34c seen in small RNA sequencing data and characterizes the targets of a miR-34c 5’ isomiR detectable by RNA sequencing.

We show that miR-34c is upregulated in ILD tissue compared to control tissue and that this dysregulation is specific to miR-34c and not other miR-34 family members miR-34a and miR-34b. These results indicate a specific role for miR-34c in ILD, especially since miR-34b and miR-34c share a promoter and are commonly transcribed together.

Previously, through the use of target prediction software, we saw that miR-34c and the miR-34c 5’ isomiR predicted targets are enriched with members of the Ras signaling gene family. Furthering these studies, we measured the predicted targets of miR-34c and the miR-34c 5’ isomiR gene expression in ILD and control tissues with a secondary platform (qRT-PCR), and confirmed that a number of these genes are in fact downregulated in ILD tissue. The large family of Ras signaling genes are important mediators of many pathways such as EGFR, PI3K, and MAPK[160]. Interestingly, ILD has been shown to have an uncommon but strong adverse reaction to EGFR inhibitors used in the treatment of lung cancer[161,162]. Together, these results potentially indicate a role for EGFR in ILD progression. A better understanding of the Ras signaling genes that are dysregulated with ILD progression could uncover the mechanism by which these EGFR inhibitors are impacting ILD.

Turning to cell biology techniques, we overexpressed miR-34c and miR-34c 5' isomiR in a fibroblast cell line and saw a significant decrease in *NOTCH1*, a positive control, with both miR-34c and miR-34c 5' isomiR expression. We also observed a significant decrease in *GRB2* expression with only the miR-34c 5' isomiR overexpression, indicating a specific role for the isomiR in regulating *GRB2*. With *GRB2* being an important upstream regulator of the EGFR pathway[163–165], the regulation of this gene by the miR-34c 5' isomiR has the potential to have important therapeutic implications if the EGFR pathway is found to be a regulator of ILD. Currently, there is research showing the EGF 61A/G polymorphism is associated with sporadic ILD[102] and increased expression of EGFR in three ILD diseases[166], leading credence to the concept that the EGFR pathway may be dysregulated in ILD.

Also of interest, miR-34a expression is significantly downregulated following overexpression of miR-34c or the miR-34c 5' isomiR. MiR-34a is well known to be a regulator of the EGFR pathway[167,168]. While the downregulation of miR-34a with overexpression of miR-34c or its variant is an unexpected result, it's possible that there is compensation in the cell between these two microRNAs to prevent overregulation of the EGFR pathway. Further studies are needed to test this hypothesis.

There appear to be mRNA targets in common and specific to miR-34c and the miR-34c 5' isomiR. This is not unlike the role different microRNA family members have, with the distinction being that instead of being produced from different areas of the chromosome, the same pre-miRNA can be cleaved to produce variants. Future studies are needed to better understand the mechanism by which isomiRs are produced during the microRNA production process. Additional studies are also needed to better understand the role of the EGFR pathway in ILD progression and the phenotypes that are regulated by the miR-34c variants in these diseases.

4.5 Conclusions

The data produced in these studies show an upregulation of miR-34c in ILD tissue compared to control. Furthermore, we have shown that miR-34c and the miR-34c 5' isomiR have shared as well as individual targets for regulation and that Ras signaling genes are regulated by the 5' miR-34c isomiR, genes that can in turn regulate the EGFR pathway. Further studies are needed to better understand the role of the EGFR pathway in ILD as well as the potential role of the miR-34c 5' isomiR in regulating this pathway as a therapeutic or therapeutic target.

CHAPTER FIVE

General Conclusions and Future Directions

The research featured in this dissertation collectively identifies microRNAs that are dysregulated in lung diseases and describes their potential role in these diseases through the use of cellular and molecular biology techniques. All the microRNA identified in this document have the potential to be utilized as therapeutics or leveraged to identify therapeutic targets in lung cancer and ILD with further follow-up studies. Specifically, the results from this dissertation assert that:

- The microRNA miR-4423 is expressed in the airway after the beginning of the differentiation of airway basal cells in a manner that correlates with ciliated cell markers and overexpression of this microRNA in the airway epithelium induces ciliogenesis. In addition, the expression of miR-4423 is lost during the progression of squamous cell carcinoma and can potentially be utilized as a cancer therapeutic in patients with this disease. These studies highlight the importance of understanding a microRNA's role in normal regulation to recognize the consequence of its dysregulation.
- miR-424 is a microRNA that is specifically upregulated in never smoker lung adenocarcinoma and potentially regulates a large number of cancer related pathways in lung adenocarcinoma. By modulating the expression of miR-424 in lung adenocarcinoma cell lines, we show that miR-424 modulates a migration phenotype in this disease. The experiments in this study stress the

need to look into personalized therapies and put forth the idea of using miR-424 as a therapeutic target in never smoker specific lung adenocarcinoma.

- MicroRNAs have many variants associated with their expression and these variants have the potential to expand the network of genes that a microRNA can regulate. In ILD, miR-34c and isomiRs of miR-34c are upregulated, potentially regulating Ras signaling genes and pathways such as EGFR to progress the disease. Additional experiments show that a 5' isomiR of miR-34c and the canonical form of miR-34c are functionally able to target independent and shared mRNA. The results of these studies bring to light the variation we see with each microRNA and caution that when we think of a microRNA as a therapeutic option, we take into account all the variants associated with that microRNA.

The continued study of microRNAs in disease may facilitate a better understanding of a microRNAs regulatory role in the cell. Advancing our knowledge of microRNAs will allow for better understanding of their potential for therapeutic applications. The work outlined in this document highlight the variable roles microRNAs have in the cell, from driving differentiation and disease to regulating numerous pathways by the formation of isomiRs, the full potential of an individual microRNA is currently still an area of exploration.

The results of these studies make a case for microRNAs playing key roles in disease biology and therefore highlight the limitless potential of using microRNAs for therapeutics. While future experiments are needed to fully understand the implications of using the microRNAs outlined in this document for therapeutic purposes, the studies in this dissertation argue for improved overall understanding of the biology of microRNAs so that we can better employ microRNAs as therapeutics. There are many unanswered questions in regards to microRNAs that hinder our knowledge on how to best use them in translational science. It is only with further research into this field that we can better utilize the potential of microRNAs in regards to treating, and even preventing disease.

APPENDIX

EDITORIAL

Liberalization of PNAS copyright policy: Noncommercial use freely allowed

We have changed our copyright and permissions policies to make it easier for authors and readers to use material published in PNAS for research or teaching. Our guiding principle is that, while PNAS retains copyright, anyone can make noncommercial use of work in PNAS without asking our permission, provided that the original source is cited. For commercial use (e.g., in books for sale or in corporate marketing materials), we approve requests on an individual basis and may ask for compensation. We have revised our copyright assignment form to make the changes clear (www.pnas.org/misc/copyright.pdf) and added to our web site a "frequently asked questions" (FAQ) section on author and reader rights (www.pnas.org/misc/authorfaq.shtml).

As a PNAS author, you automatically have the right to do the following:

- Post a PDF of your article on your web site.
- Post a webcast containing material from your article.
- Make electronic or hard copies of articles for your personal use, including classroom use.
- Use, after publication, all or part of your article in a printed compilation of your work, such as collected writings or lecture notes.
- Include your article in your thesis or dissertation.
- Reuse your original figures or tables in your future works.
- Post a preprint of your article on a public electronic server, provided that you do not use the files created by PNAS.
- Present your paper at a meeting or conference, including those that are webcast, and give copies of your paper to meeting attendees before or after publication in PNAS. For interactions with the media prior to publication, see the PNAS policy on media coverage (www.pnas.org/misc/forms.shtml).
- Permit others to use your original figures or tables published in PNAS for noncommercial use (e.g., in a review article), provided that the source is cited. Third parties need not request permission to use figures and tables for such use.

Given that authors and readers can automatically use original material in PNAS for research or teaching, why do we request copyright transfer? We do so for three reasons: to allow us to publish, archive, and migrate articles to new media; to remove the administrative burden of rights and permissions management from authors; and to provide protection from copyright abuse.

We do not feel that this or any copyright policy is the only one possible. In fact, our policy has changed through our 90 years of publishing and surely will change again. We have requested that authors transfer copyright only since 1993. From the first issue of PNAS in 1915 through 1992, authors held copyright to their articles. From 1978 to 1992, we registered copyright for each journal issue as a collected work but did not request copyright for individual articles. In 1993, we began requiring that authors transfer copyright "in all forms, languages, and media now or hereafter known," which granted us the rights to publish papers online in 1997 and to then digitize selected back issues and post them online.

We think that our current policy best meets the needs of readers, authors, and the journal, for the following reasons:

1. **To store and migrate archival formats of the journal.** We are committed to facilitating permanent, freely accessible archives of the scientific literature. PNAS is a charter member of PubMed Central, a digital archive of the life sciences journal literature (www.pubmedcentral.nih.gov), and is a participant in the National Library of Medicine's effort to digitize and post back issues of journals. Not holding copyright to individual articles from 1915 to 1992 delayed our posting of this older material online because we do not have the legal rights to do so. In the end we proceeded without explicit permission from the original authors or their heirs. We accept the risk in doing so because we believe it is clearly in everyone's best interest. If a copyright holder objects, however, we will immediately remove the article from our online collection. Full copyright transfer allows publishers explicit rights to invest in long-term archiving strategies.
2. **To provide an administrative convenience for everyone.** Despite our liberal rights and permissions policies, PNAS still receives more than 50 commercial and noncommercial permission requests per week. We routinely agree to noncommercial use, so such requests waste everyone's time.

   Unfortunately, PNAS cannot provide permission for others to use all or part of articles published from 1915 to 1992 because we do not hold copyright. Only the original authors or their designees can grant permission. Researchers are frustrated when they contact us for permission to use seminal works and we are unable to grant their requests.
3. **To provide international protection regarding infringement or plagiarism.** On the rare occasion that material is misused, authors appeal to PNAS to intervene on their behalf to enforce copyright protection. In such cases, a formal query from PNAS or the threat of a copyright infringement lawsuit has prompted expeditious action. In cases of redundant publication we sanction authors for violating journal and copyright policy. Because international standards and copyright law are complex, PNAS leaves interpretation of global copyright standards to our expert legal counsel.

We also support creative efforts such as charting, mining, analyzing, sorting, navigating, and displaying information contained in PNAS. The highly successful Sackler Colloquium "Mapping Knowledge Domains" (www.pnas.org/content/vol101/suppl_1) is a prime example (1). We encourage authors to use standard forms of data presentation to facilitate this process.

Nicholas R. Cozzarelli,
Editor-in-Chief

Kenneth R. Fulton,
Publisher

Diane M. Sullenberger,
Executive Editor

1. Shiffrin, R. M. & Börner, K., eds. (2004) *Proc. Natl. Acad. Sci. USA* **101**, Suppl. 1, 5183–5310.

www.pnas.org/cgi/doi/10.1073/pnas.0405514101

PNAS | **August 24, 2004** | vol. 101 | no. 34 | **12399**

LIST OF JOURNAL ABBREVIATIONS

Adv. Exp. Med. Biol.	Advances in Experimental Medicine and Biology
Am. J. Pathol.	American Journal of Pathology
Am. J. Physiol.	American Journal of Physiology
Am. J. Respir. Cell Mol. Biol.	American Journal of Respiratory Cell and Molecular Biology
Am. J. Respir. Crit. Care Med.	American Journal of Respiratory and Critical Care Medicine
Annu. Rev. Med.	Annual Review of Medicine
Annu. Rev. Pharmacol. Toxicol.	Annual Review of Pharmacology and Toxicology
Arch. Bronconeumol.	Archivos de Bronconeumología
Biochim. Biophys. Acta	Biochimica et Biophysica Acta
BioMed Res. Int.	BioMed Research International
Br. J. Cancer	British Journal of Cancer
CA Cancer J. Clin.	CA: A Cancer Journal for Clinicians
Cancer Lett.	Cancer Letters
Cancer Prev. Res. (Phila)	Cancer Prevention Research (Philadelphia, PA.)
Cancer Res.	Cancer Research
Cell Res.	Cell Research
Cell. Mol. Life Sci.	Cellular and Molecular Life Science
Cell. Physiol. Biochem.	Cellular Physiology And Biochemistry
Chin. Med. J. (Engl.)	Chinese Medical Journal (English Edition)

Clin. Biochem.	Clinical Biochemistry
Clin. Cancer Res.	Clinical Cancer Research
Curr. Opin. Cell Biol.	Current Opinion in Cell Biology
Curr. Protoc. Pharmacol.	Current Protocols in Pharmacology
Dis. Model. Mech.	Disease Models & Mechanisms
EMBO J.	EMBO Journal
Eur. J. Cancer	European Journal of Cancer
Eur. J. Pharmacol.	European Journal of Pharmacology
Eur. Respir. Rev.	European Respiratory Review
Exp. Ther. Med.	Experimental and Therapeutic Medicine
Gene Ther.	Gene Therapy
Genet. Res. Int.	Genetics Research International
Genome Biol.	Genome Biology
Gynecol. Oncol.	Gynecologic Oncology
Hum. Mol. Genet.	Human Molecular Genetics
In Vitro Cell. Dev. Biol. Anim.	In Vitro Cellular & Developmental Biology - Animal
Int. J. Biochem. Cell Biol.	The International Journal of Biochemistry & Cell Biology
Int. J. Cancer	International Journal of Cancer
Int. J. Clin. Exp. Pathol.	International Journal of Clinical and Experimental Pathology
Int. J. Mol. Sci.	International Journal of Molecular Sciences
Int. J. Oncol.	International Journal of Oncology
J. Biol. Chem.	Journal of Biological Chemistry

J. Bone Miner. Res.	Journal of Bone Mineral Research
J. Bras. Pneumol.	Jornal Brasileiro de Pneumologia
J. Carcinog.	Journal of Carcinogenesis
J. Cell Sci.	Journal of Cell Science
J. Chemother.	Journal of Chemotherapy
J. Clin. Invest.	Journal of Clinical Investigation
J. Clin. Oncol.	Journal of Clinical Oncology
J. Control. Release	Journal of Controlled Release
J. Dig. Dis.	Journal of Digestive Diseases
J. Mol. Cell Biol.	Journal of Molecular Cell Biology
J. Natl. Cancer Inst.	Journal of the National Cancer Institute
J. Thorac. Oncol.	Journal of Thoracic Oncology
Lancet Oncol.	Lancet Oncology
Lancet Respir. Med.	Lancet Respiratory Medicine
Liver Int.	Liver International
Mayo Clin. Proc.	Mayo Clinic Proceedings
Mol. Biol. Cell	Molecular Biology of the Cell
Mol. Cancer Res.	Molecular Cancer Research
Mol. Cell. Biol.	Molecular and Cellular Biology
Mol. Oncol.	Molecular Oncology
Mol. Ther. Nucleic Acids	Molecular Therapy - Nucleic Acids
N. Engl. J. Med.	The New England Journal of Medicine
Nat. Cell Biol.	Nature Cell Biology
Nat. Med.	Nature Medicine

Nat. Rev. Cancer	Nature Reviews Cancer
Nat. Rev. Drug Discov.	Nature Reviews Drug Discovery
Nat. Rev. Genet.	Nature Reviews Genetics
Nat. Rev. Mol. Cell Biol.	Nature Reviews Molecular Cell Biology
Nucleic Acids Res.	Nucleic Acids Research
Open Biol.	Open Biology
Paediatr. Respir. Rev.	Paediatric Respiratory Reviews
Pathol. Oncol. Res.	Pathology and Oncology Research
PLoS Med.	PLOS Medicine
Proc. Am. Thorac. Soc.	Proceedings of the American Thoracic Society
Proc. Natl. Acad. Sci. U.S.A.	Proceedings of the National Academy of Sciences of the United States of America
Radiol. Clin. North Am.	Radiologic Clinics of North America
Respir. Physiol. Neurobiol.	Respiratory Physiology & Neurobiology
Respir. Res.	Respiratory Research
Risk Anal.	Risk Analysis
Sci. Rep.	Scientific Reports
Sci. Transl. Med.	Science Translational Medicine
Semin. Roentgenol.	Seminars in Roentgenology
Transl. Res.	Translational Research
Trends Genet.	Trends in Genetics
Trends Mol. Med.	Trends in Molecular Medicine
World J. Gastroenterol.	World Journal of Gastroenterology

BIBLIOGRAPHY

1. Rackley, C. R. & Stripp, B. R. Building and maintaining the epithelium of the lung. *J. Clin. Invest.* **122,** 2724–2730 (2012).

2. Yahaya, B. Understanding cellular mechanisms underlying airway epithelial repair: selecting the most appropriate animal models. *ScientificWorldJournal* **2012,** 961684 (2012).

3. Vaughan, A. E. & Chapman, H. A. Regenerative activity of the lung after epithelial injury. *Biochim. Biophys. Acta* **1832,** 922-930 (2012).

4. Lambrecht, B. N. & Hammad, H. Death at the airway epithelium in asthma. *Cell Res.* **23,** 588-589 (2013).

5. Rock, J. R., Randell, S. H. & Hogan, B. L. M. Airway basal stem cells: a perspective on their roles in epithelial homeostasis and remodeling. *Dis. Model. Mech.* **3,** 545–556 (2010).

6. Gharib, S. A. *et al.* Matrix metalloproteinase-7 coordinates airway epithelial injury response and differentiation of ciliated cells. *Am. J. Respir. Cell Mol. Biol.* **48,** 390–396 (2013).

7. Eltom, S., Stevenson, C. & Birrell, M. A. Cigarette Smoke Exposure as a Model of Inflammation Associated with COPD. *Curr. Protoc. Pharmacol. Editor. Board SJ Enna Ed.--Chief Al* **Chapter 5,** Unit5.64 (2013).

8. Kuchar, E. *et al.* Immunopathology of exercise-induced bronchoconstriction in athletes - a new modified inflammatory hypothesis. *Respir. Physiol. Neurobiol.* **187,** 82-87 (2013).

9. Siva, R., Birring, S. S., Berry, M., Rowbottom, A. & Pavord, I. D. Peptic ulceration, Helicobacter Pylori seropositivity and Chronic Obstructive Pulmonary Disease. *Respirology* **18,** 728-731 (2013).

10. Juncadella, I. J. *et al.* Apoptotic cell clearance by bronchial epithelial cells critically influences airway inflammation. *Nature* **493,** 547–551 (2013).

11. Gregory, L. G. *et al.* IL-25 drives remodelling in allergic airways disease induced by house dust mite. *Thorax* **68,** 82–90 (2013).

12. Solberg, O. D. *et al.* Airway epithelial miRNA expression is altered in asthma. *Am. J. Respir. Crit. Care Med.* **186,** 965–974 (2012).

13. Zhai, Y. *et al.* Coordinated changes in mRNA turnover, translation, and RNA processing bodies in bronchial epithelial cells following inflammatory stimulation. *Mol. Cell. Biol.* **28,** 7414–7426 (2008).

14. Wang, Z. Association between chronic obstructive pulmonary disease and lung cancer: the missing link. *Chin. Med. J. (Engl.)* **126,** 154–165 (2013).

15. Fujimoto, J. *et al.* G-protein coupled receptor family C, group 5, member A (GPRC5A) expression is decreased in the adjacent field and normal bronchial epithelia of patients with chronic obstructive pulmonary disease and non-small-cell lung cancer. *J. Thorac. Oncol. Off. Publ. Int. Assoc. Study Lung Cancer* **7,** 1747–1754 (2012).

16. Cruse, G. & Bradding, P. Mast cells in airway diseases and interstitial lung disease. *Eur. J. Pharmacol.* Advance online publication. doi:10.1016/j.ejphar.2015.04.046 (2015).

17. Johannson, K. A., Balmes, J. R. & Collard, H. R. Air pollution exposure: a novel environmental risk factor for interstitial lung disease? *Chest* **147,** 1161–1167 (2015).

18. Chronic Inflammation and Cancer | Cancer Network. (2002). Accessed 2 Oct 2015. at <http://www.cancernetwork.com/review-article/chronic-inflammation-and-cancer>

19. Brody, J. S. Transcriptome alterations induced by cigarette smoke. *Int. J. Cancer J. Int. Cancer* **131,** 2754–2762 (2012).

20. Steiling, K., Lenburg, M. E. & Spira, A. Airway gene expression in chronic obstructive pulmonary disease. *Proc. Am. Thorac. Soc.* **6,** 697–700 (2009).

21. Beane, J. *et al.* Characterizing the impact of smoking and lung cancer on the airway transcriptome using RNA-Seq. *Cancer Prev. Res. Phila. Pa* **4,** 803–817 (2011).

22. Gower, A. C., Steiling, K., Brothers, J. F., Lenburg, M. E. & Spira, A. Transcriptomic Studies of the Airway Field of Injury Associated with Smoking-Related Lung Disease. *Proc. Am. Thorac. Soc.* **8,** 173–179 (2011).

23. Beane, J. *et al.* A prediction model for lung cancer diagnosis that integrates genomic and clinical features. *Cancer Prev. Res. Phila. Pa* **1,** 56–64 (2008).

24. Gustafson, A. M. *et al.* Airway PI3K pathway activation is an early and reversible event in lung cancer development. *Sci. Transl. Med.* **2,** 26ra25 (2010).

25. Spira, A. *et al.* Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. *Nat. Med.* **13,** 361–366 (2007).

26. Kadara, H. *et al.* Characterizing the molecular spatial and temporal field of injury in early-stage smoker non-small cell lung cancer patients after definitive surgery by expression profiling. *Cancer Prev. Res. Phila. Pa* **6,** 8–17 (2013).

27. Steiling, K. *et al.* A Dynamic Bronchial Airway Gene Expression Signature of COPD and Lung Function Impairment. *Am. J. Respir. Crit. Care Med.* **187,** 933-942 (2013).

28. Beane, J. *et al.* SIRT1 pathway dysregulation in the smoke-exposed airway epithelium and lung tumor tissue. *Cancer Res.* **72,** 5702–5711 (2012).

29. Siegel, R., Naishadham, D. & Jemal, A. Cancer statistics, 2013. *CA. Cancer J. Clin.* **63,** 11–30 (2013).

30. McErlean, A. & Ginsberg, M. S. Epidemiology of Lung Cancer. *Semin. Roentgenol.* **46,** 173–177 (2011).

31. Youlden, D. R., Cramb, S. M. & Baade, P. D. The International Epidemiology of Lung Cancer: geographical distribution and secular trends. *J. Thorac. Oncol. Off. Publ. Int. Assoc. Study Lung Cancer* **3,** 819–831 (2008).

32. Novaes, F. T. *et al.* Lung cancer: histology, staging, treatment and survival. *J. Bras. Pneumol.* **34,** 595–600 (2008).

33. Lung cancer - small cell: MedlinePlus Medical Encyclopedia. Accessed 2 Oct 2015. at <https://www.nlm.nih.gov/medlineplus/ency/article/000122.htm>

34. Ginsberg, M. S., Grewal, R. K. & Heelan, R. T. Lung Cancer. *Radiol. Clin. North Am.* **45,** 21–43 (2007).

35. Sutherland, K. D. & Berns, A. Cell of origin of lung cancer. *Mol. Oncol.* **4,** 397–403 (2010).

36. Onaitis, M. & Hanna, J. Cell of origin of lung cancer. *J. Carcinog.* **12,** 6 (2013).

37. McCarthy, W. J., Meza, R., Jeon, J. & Moolgavkar, S. H. Chapter 6: Lung Cancer in Never Smokers: Epidemiology and Risk Prediction Models. *Risk Anal.* **32,** S69–S84 (2012).

38. Samet, J. M. *et al.* Lung Cancer In Never Smokers: Clinical Epidemiology And Environmental Risk Factors. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **15,** 5626–5645 (2009).

39. Couraud, S., Zalcman, G., Milleron, B., Morin, F. & Souquet, P.-J. Lung cancer in never smokers--a review. *Eur. J. Cancer Oxf. Engl. 1990* **48,** 1299–1311 (2012).

40. Sun, S., Schiller, J. H. & Gazdar, A. F. Lung cancer in never smokers — a different disease. *Nat. Rev. Cancer* **7,** 778–790 (2007).

41. Wu, K., House, L., Liu, W. & Cho, W. C. S. Personalized Targeted Therapy for Lung Cancer. *Int. J. Mol. Sci.* **13,** 11471–11496 (2012).

42. Govindan, R. *et al.* Genomic Landscape Of Non-Small Cell Lung Cancer In Smokers And Never Smokers. *Cell* **150,** 1121–1134 (2012).

43. Santoro, I. L., Ramos, R. P., Franceschini, J., Jamnik, S. & Fernandes, A. L. G. Non-small cell lung cancer in never smokers: a clinical entity to be identified. *Clinics* **66,** 1873–1877 (2011).

44. Antoniou, K. M. *et al.* Interstitial lung disease. *Eur. Respir. Rev.* **23,** 40–54 (2014).

45. Nathan, N., Corvol, H., Amselem, S. & Clement, A. Biomarkers in Interstitial lung diseases. *Paediatr. Respir. Rev.* Advance online Publication. doi:10.1016/j.prrv.2015.05.002 (2015).

46. Steele, M. P. & Schwartz, D. A. Molecular Mechanisms in Progressive Idiopathic Pulmonary Fibrosis. *Annu. Rev. Med.* **64,** 265–276 (2013).

47. Guillot, L. *et al.* Alveolar epithelial cells: Master regulators of lung homeostasis. *Int. J. Biochem. Cell Biol.* **45,** 2568–2573 (2013).

48. Bjoraker, J. A. *et al.* Prognostic significance of histopathologic subsets in idiopathic pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* **157,** 199–203 (1998).

49. Carrington, C. B., Gaensler, E. A., Coutu, R. E., FitzGerald, M. X. & Gupta, R. G. Natural history and treated course of usual and desquamative interstitial pneumonia. *N. Engl. J. Med.* **298,** 801–809 (1978).

50. Stack, B. H., Choo-Kang, Y. F. & Heard, B. E. The prognosis of cryptogenic fibrosing alveolitis. *Thorax* **27,** 535–542 (1972).

51. Lee, R. C., Feinbaum, R. L. & Ambros, V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* **75,** 843–854 (1993).

52. Lau, N. C., Lim, L. P., Weinstein, E. G. & Bartel, D. P. An abundant class of tiny RNAs with probable regulatory roles in Caenorhabditis elegans. *Science* **294,** 858–862 (2001).

53. Oglesby, I. K., McElvaney, N. G. & Greene, C. M. MicroRNAs in inflammatory lung disease--master regulators or target practice? *Respir. Res.* **11,** 148 (2010).

54. Lee, Y. *et al.* MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.* **23,** 4051–4060 (2004).

55. Bohnsack, M. T., Czaplinski, K. & Gorlich, D. Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *RNA N. Y. N* **10,** 185–191 (2004).

56. Kim, V. N. MicroRNA biogenesis: coordinated cropping and dicing. *Nat. Rev. Mol. Cell Biol.* **6,** 376–385 (2005).

57. Weinmann, L. *et al.* Importin 8 is a gene silencing factor that targets argonaute proteins to distinct mRNAs. *Cell* **136,** 496–507 (2009).

58. Liu, J. Control of protein synthesis and mRNA degradation by microRNAs. *Curr. Opin. Cell Biol.* **20,** 214–221 (2008).

59. Sittka, A. & Schmeck, B. MicroRNAs in the Lung. *Adv. Exp. Med. Biol.* **774,** 121–134 (2013).

60. Mach, C. *et al.* Novel MicroRNAs regulating proliferation and apoptosis in uterine papillary serous carcinomas. *Cancer Lett.* **335,** 314-322 (2013).

61. Tétreault, N. & De Guire, V. miRNAs: Their discovery, biogenesis and mechanism of action. *Clin. Biochem.* **46,** 842-845 (2013).

62. Wang, L.-G. *et al.* MicroRNA-34b functions as a tumor suppressor and acts as a nodal point in the feedback loop with Met. *Int. J. Oncol.* **42,** 957–962 (2013).

63. Tang, T. *et al.* MicroRNA-182 plays an onco-miRNA role in cervical cancer. *Gynecol. Oncol.* **129,** 199-208 (2013).

64. Carroll, A. P., Tooney, P. A. & Cairns, M. J. Context-specific microRNA function in developmental complexity. *J. Mol. Cell Biol.* **5,** 73-84 (2013).

65. Zhang, X., Graves, P. & Zeng, Y. Overexpression of human Argonaute2 inhibits cell and tumor growth. *Biochim. Biophys. Acta* **1830,** 2553–2561 (2012).

66. Enfield, K. S. S., Pikor, L. A., Martinez, V. D. & Lam, W. L. Mechanistic Roles of Noncoding RNAs in Lung Cancer Biology and Their Clinical Implications. *Genet. Res. Int.* **2012,** 737416 (2012).

67. Williams, A. E. Functional aspects of animal microRNAs. *Cell. Mol. Life Sci. CMLS* **65,** 545–562 (2008).

68. Angulo, M., Lecuona, E. & Sznajder, J. I. Role of MicroRNAs in Lung Disease. *Arch. Bronconeumol.* **48,** 325-330 (2012).

69. Kang, J. *et al.* microRNA-99b acts as a tumor suppressor in non-small cell lung cancer by directly targeting fibroblast growth factor receptor 3. *Exp. Ther. Med.* **3,** 149–153 (2012).

70. Lee, J. M. *et al.* The novel miR-7515 decreases the proliferation and migration of human lung cancer cells by targeting c-Met. *Mol. Cancer Res. MCR* **11,** 43–53 (2013).

71. Zhang, H. *et al.* MiR-138 Inhibits Tumor Growth Through Repression of EZH2 in Non-Small Cell Lung Cancer. *Cell. Physiol. Biochem. Int. J. Exp. Cell. Physiol. Biochem. Pharmacol.* **31,** 56–65 (2013).

72. Xu, L. *et al.* MicroRNA-7-regulated TLR9 signaling-enhanced growth and metastatic potential of human lung cancer cells by altering the phosphoinositide-3-kinase, regulatory subunit 3/Akt pathway. *Mol. Biol. Cell* **24,** 42–55 (2013).

73. Pandit, K. V., Milosevic, J. & Kaminski, N. MicroRNAs in idiopathic pulmonary fibrosis. *Transl. Res.* **157,** 191–199 (2011).

74. Sendi, H. *et al.* MiR-122 decreases HCV entry into hepatocytes through binding to the 3' UTR of OCLN mRNA. *Liver Int. Off. J. Int. Assoc. Study Liver* **35,** 1315–1323 (2015).

75. Bader, A. G., Brown, D., Stoudemire, J. & Lammers, P. Developing therapeutic microRNAs for cancer. *Gene Ther.* **18,** 1121–1126 (2011).

76. Ha, M. & Kim, V. N. Regulation of microRNA biogenesis. *Nat. Rev. Mol. Cell Biol.* **15,** 509–524 (2014).

77. Yin, H. *et al.* Non-viral vectors for gene-based therapy. *Nat. Rev. Genet.* **15,** 541–555 (2014).

78. Li, Z., Yang, C.-S., Nakashima, K. & Rana, T. M. Small RNA-mediated regulation of iPS cell generation. *EMBO J.* **30,** 823–834 (2011).

79. Bennett, C. F. & Swayze, E. E. RNA targeting therapeutics: molecular mechanisms of antisense oligonucleotides as a therapeutic platform. *Annu. Rev. Pharmacol. Toxicol.* **50,** 259–293 (2010).

80. Lam, J. K. W., Chow, M. Y. T., Zhang, Y. & Leung, S. W. S. siRNA Versus miRNA as Therapeutics for Gene Silencing. *Mol. Ther. Nucleic Acids* **4,** e252 (2015).

81. Wen, D., Danquah, M., Chaudhary, A. K. & Mahato, R. I. Small molecules targeting microRNA for cancer therapy: Promises and obstacles. *J. Control. Release* Advanced online publication. doi:10.1016/j.jconrel.2015.08.011 (2015)

82. Mello Biotech - Innovative microRNA Therapeutics. Accessed 5 Oct 2015. at <http://www.mellobiotech.com/>

83. Regulus Therapeutics – A Leading microRNA Therapeutics Company | Regulus Therapeutics – a leading company developing microRNA therapeutics. Medicines that target the pathways of human disease. Accessed 5 Oct 2015. at <http://www.regulusrx.com/>

84. miRagen Therapeutics | microRNA Based Therapeutics. Accessed 5 Oct 2015. at <http://miragentherapeutics.com/>

85. InteRNA Technologies - Intelligent miRNA therapeutics for cancer. Accessed 5 Oct 2015. at <www.interna-technologies.com>

86. Mirna Therapeutics - Using microRNAs to resore the body's natural ability to fight cancer. Accessed 5 Oct 2015. at <www.mirrnarx.com>

87. Groove Biopharma. Accessed 5 Oct 2015. at <www.groovebiopharma.com>

88. Hayes, J., Peruzzi, P. P. & Lawler, S. MicroRNAs in cancer: biomarkers, functions and therapy. *Trends Mol. Med.* **20,** 460–469 (2014).

89. Hydbring, P. & Badalian-Very, G. Clinical applications of microRNAs. *F1000Research* **2,** (2013).

90. MRX34. Accessed 5 Oct 2015. at <www.mirnarx.com/pipeline/mirna-MRX34.html>

91. A Multicenter Phase I Study of MRX34, MicroRNA miR-RX34 Liposomal Injection. Accessed 5 Oct 2015. at <https://www.clinicaltrials.gov/ct2/show/NCT01829971?term=mirna+therapeutics&rank=115>

92. Janssen, H. L. A. *et al.* Treatment of HCV infection by targeting microRNA. *N. Engl. J. Med.* **368,** 1685–1694 (2013).

93. Asselin-Labat, M.-L. & Filby, C. E. Adult lung stem cells and their contribution to lung tumourigenesis. *Open Biol.* **2,** 120094 (2012).

94. Marcet, B. *et al.* Control of vertebrate multiciliogenesis by miR-449 through direct repression of the Delta/Notch pathway. *Nat. Cell Biol.* **13,** 693–699 (2011).

95. Luo, W. *et al.* MicroRNA-449a is downregulated in non-small cell lung cancer and inhibits migration and invasion by targeting c-Met. *PloS One* **8,** e64759 (2013).

96. Perdomo, C. *et al.* MicroRNA 4423 is a primate-specific regulator of airway epithelial cell differentiation and lung carcinogenesis. *Proc. Natl. Acad. Sci.* **110,** 18946-18951 (2013).

97. Jemal, A. *et al.* Global cancer statistics. *CA. Cancer J. Clin.* **61,** 69–90 (2011).

98. Wang, Y. X. *et al.* Initial study of microRNA expression profiles of colonic cancer without lymph node metastasis. *J. Dig. Dis.* **11,** 50–54 (2010).

99. Wu, K. *et al.* MicroRNA-424-5p Suppresses the Expression of SOCS6 in Pancreatic Cancer. *Pathol. Oncol. Res.* **19,** 739–748 (2013).

100. Spagnolo, P., Grunewald, J. & du Bois, R. M. Genetic determinants of pulmonary fibrosis: evolving concepts. *Lancet Respir. Med.* **2,** 416–428 (2014).

101. Qi, W.-X., Sun, Y.-J., Shen, Z. & Yao, Y. Risk of interstitial lung disease associated with EGFR-TKIs in advanced non-small-cell lung cancer: a meta-analysis of 24 phase III clinical trials. *J. Chemother. Florence Italy* **27,** 40–51 (2015).

102. Li, C. *et al.* Epidermal Growth Factor Receptor (EGFR) Pathway Genes and Interstitial Lung Disease: An Association Study. *Sci. Rep.* **4,** 4893 (2014).

103. Basbaum, C. & Jany, B. Plasticity in the airway epithelium. *Am. J. Physiol.* **259,** L38–46 (1990).

104. Antoshina, E. & Ostrowski, L. E. TGF beta 1 induces growth arrest and apoptosis but not ciliated cell differentiation in rat tracheal epithelial cell cultures. *In Vitro Cell. Dev. Biol. Anim.* **33,** 212–217 (1997).

105. Randell, S. H. Progenitor-progeny relationships in airway epithelium. *Chest* **101,** 11S–16S (1992).

106. Marcet, B., Chevalier, B., Coraux, C., Kodjabachian, L. & Barbry, P. MicroRNA-based silencing of Delta/Notch signaling promotes multiple cilia formation. *Cell Cycle Georget. Tex* **10,** 2858–2864 (2011).

107. Lizé, M., Herr, C., Klimke, A., Bals, R. & Dobbelstein, M. MicroRNA-449a levels increase by several orders of magnitude during mucociliary differentiation of airway epithelia. *Cell Cycle Georget. Tex* **9,** 4579–4583 (2010).

108. Wu, H. & Leng, R. P. UBE4B, a ubiquitin chain assembly factor, is required for MDM2-mediated p53 polyubiquitination and degradation. *Cell Cycle Georget. Tex* **10,** 1912–1915 (2011).

109. Wu, H. *et al.* UBE4B promotes Hdm2-mediated degradation of the tumor suppressor p53. *Nat. Med.* **17,** 347–355 (2011).

110. Zhang, Y., Lv, Y., Zhang, Y. & Gao, H. Regulation of p53 level by UBE4B in breast cancer. *PloS One* **9,** e90154 (2014).

111. Alexopoulou, D. K., Papadopoulos, I. N. & Scorilas, A. Clinical significance of kallikrein-related peptidase (KLK10) mRNA expression in colorectal cancer. *Clin. Biochem.* **46,** 1453–1461 (2013).

112. Jiao, X. *et al.* Overexpression of kallikrein gene 10 is a biomarker for predicting poor prognosis in gastric cancer. *World J. Gastroenterol. WJG* **19,** 9425–9431 (2013).

113. Jiang, R., Shi, Z., Johnson, J. J., Liu, Y. & Stack, M. S. Kallikrein-5 Promotes Cleavage of Desmoglein-1 and Loss of Cell-Cell Cohesion in Oral Squamous Cell Carcinoma. *J. Biol. Chem.* **286,** 9127–9135 (2011).

114. Johnson, S. K., Ramani, V. C., Hennings, L. & Haun, R. S. Kallikrein 7 enhances pancreatic cancer cell invasion by shedding E-cadherin. *Cancer* **109,** 1811–1820 (2007).

115. Nava, P., Kamekura, R. & Nusrat, A. Cleavage of transmembrane junction proteins and their role in regulating epithelial homeostasis. *Tissue Barriers* **1,** e24783 (2013).

116. Zhang, J. *et al.* LRP8 mediates Wnt/β-catenin signaling and controls osteoblast differentiation. *J. Bone Miner. Res. Off. J. Am. Soc. Bone Miner. Res.* **27,** 2065–2074 (2012).

117. Pencheva, N. *et al.* Convergent Multi-miRNA Targeting of ApoE Drives LRP1/LRP8-Dependent Melanoma Metastasis and Angiogenesis. *Cell* **151,** 1068–1082 (2012).

118. Garnis, C. *et al.* Involvement of multiple developmental genes on chromosome 1p in lung tumorigenesis. *Hum. Mol. Genet.* **14,** 475–482 (2005).

119. Lizé, M., Klimke, A. & Dobbelstein, M. MicroRNA-449 in cell fate determination. *Cell Cycle Georget. Tex* **10,** 2874–2882 (2011).

120. Tammemägi, M. C. *et al.* Evaluation of the lung cancer risks at which to screen ever- and never-smokers: screening rules applied to the PLCO and NLST cohorts. *PLoS Med.* **11,** e1001764 (2014).

121. Lee, P. N. & Forey, B. A. Indirectly estimated absolute lung cancer mortality rates by smoking status and histological type based on a systematic review. *BMC Cancer* **13,** 189 (2013).

122. Wakelee, H. A. *et al.* Lung cancer incidence in never smokers. *J. Clin. Oncol. Off. J. Am. Soc. Clin. Oncol.* **25,** 472–478 (2007).

123. Lim, S.-T. *et al.* Gefitinib is more effective in never-smokers with non-small-cell lung cancer: experience among Asian patients. *Br. J. Cancer* **93,** 23–28 (2005).

124. Mok, T. S. *et al.* Gefitinib or carboplatin-paclitaxel in pulmonary adenocarcinoma. *N. Engl. J. Med.* **361,** 947–957 (2009).

125. Kobayashi, S. *et al.* EGFR mutation and resistance of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* **352,** 786–792 (2005).

126. Imielinski, M. *et al.* Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* **150,** 1107–1120 (2012).

127. Subramanian, J. & Govindan, R. Molecular genetics of lung cancer in people who have never smoked. *Lancet Oncol.* **9,** 676–682 (2008).

128. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511,** 543–550 (2014).

129. Huang, G. T., Athanassiou, C. & Benos, P. V. mirConnX: condition-specific mRNA-microRNA network integrator. *Nucleic Acids Res.* **39,** W416–423 (2011).

130. Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14,** 128 (2013).

131. Agarwal, V., Bell, G. W., Nam, J.-W. & Bartel, D. P. Predicting effective microRNA target sites in mammalian mRNAs. *eLife* **4,** e05005 (2015).

132. Simões-Correia, J. *et al.* DNAJB4 molecular chaperone distinguishes WT from mutant E-cadherin, determining their fate in vitro and in vivo. *Hum. Mol. Genet.* **23,** 2094-2105 (2013).

133. Frullanti, E. *et al.* Association of lung adenocarcinoma clinical stage with gene expression pattern in noninvolved lung tissue. *Int. J. Cancer* **131,** E643–E648 (2012).

134. Ou, D., Yang, H., Hua, D., Xiao, S. & Yang, L. Novel roles of TMEM100: inhibition metastasis and proliferation of hepatocellular carcinoma. *Oncotarget* **6,** 17379–17390 (2015).

135. Garzon, R., Marcucci, G. & Croce, C. M. Targeting microRNAs in cancer: rationale, strategies and challenges. *Nat. Rev. Drug Discov.* **9,** 775–789 (2010).

136. Yang, N. *et al.* MicroRNA Microarray Identifies Let-7i as a Novel Biomarker and Therapeutic Target in Human Epithelial Ovarian Cancer. *Cancer Res.* **68,** 10307–10314 (2008).

137. Kadara, H. *et al.* Transcriptomic architecture of the adjacent airway field cancerization in non-small cell lung cancer. *J. Natl. Cancer Inst.* **106,** dju004 (2014).

138. Spira, A. *et al.* Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. *Nat. Med.* **13,** 361–366 (2007).

139. Silvestri, G. A. *et al.* A Bronchial Genomic Classifier for the Diagnostic Evaluation of Lung Cancer. *N. Engl. J. Med.* **373,** 243–251 (2015).

140. Tang, X. *et al.* EGFR Tyrosine Kinase Domain Mutations Are Detected in Histologically Normal Respiratory Epithelium in Lung Cancer Patients. *Cancer Res.* **65,** 7568–7572 (2005).

141. June, 21 & 2005. Interstitial lung disease: clinical features and management. Accessed 24 Sept 2015. at <http://www.nursingtimes.net/interstitial-lung-disease-clinical-features-and-management/203835.article>

142. Ryu, J. H. *et al.* Idiopathic pulmonary fibrosis: evolving concepts. *Mayo Clin. Proc.* **89,** 1130–1142 (2014).

143. American Thoracic Society. Idiopathic pulmonary fibrosis: diagnosis and treatment. International consensus statement. American Thoracic Society (ATS), and the European Respiratory Society (ERS). *Am. J. Respir. Crit. Care Med.* **161,** 646–664 (2000).

144. Flaherty, K. R. *et al.* Histopathologic variability in usual and nonspecific interstitial pneumonias. *Am. J. Respir. Crit. Care Med.* **164,** 1722–1727 (2001).

145. Nicholson, A. G., Colby, T. V., du Bois, R. M., Hansell, D. M. & Wells, A. U. The prognostic significance of the histologic pattern of interstitial pneumonia in patients presenting with the clinical entity of cryptogenic fibrosing alveolitis. *Am. J. Respir. Crit. Care Med.* **162,** 2213–2217 (2000).

146. Selman, M. & Pardo, A. Role of epithelial cells in idiopathic pulmonary fibrosis: from innocent targets to serial killers. *Proc. Am. Thorac. Soc.* **3,** 364–372 (2006).

147. Kaminski, N. *et al.* Global analysis of gene expression in pulmonary fibrosis reveals distinct programs regulating lung inflammation and fibrosis. *Proc. Natl. Acad. Sci.* **97,** 1778–1783 (2000).

148. Chilosi, M. *et al.* Aberrant Wnt/beta-catenin pathway activation in idiopathic pulmonary fibrosis. *Am. J. Pathol.* **162,** 1495–1502 (2003).

149. Selman, M., Pardo, A. & Kaminski, N. Idiopathic Pulmonary Fibrosis: Aberrant Recapitulation of Developmental Programs? *PLoS Med* **5,** e62 (2008).

150. Pittet, J. F. *et al.* TGF-beta is a critical mediator of acute lung injury. *J. Clin. Invest.* **107,** 1537–1544 (2001).

151. Seibold, M. A. *et al.* A common MUC5B promoter polymorphism and pulmonary fibrosis. *N. Engl. J. Med.* **364,** 1503–1512 (2011).

152. Milosevic, J. *et al.* Profibrotic role of miR-154 in pulmonary fibrosis. *Am. J. Respir. Cell Mol. Biol.* **47,** 879–887 (2012).

153. Neilsen, C. T., Goodall, G. J. & Bracken, C. P. IsomiRs – the overlooked repertoire in the dynamic microRNAome. *Trends Genet.* **28,** 544–549 (2012).

154. Cloonan, N. *et al.* MicroRNAs and their isomiRs function cooperatively to target common biological pathways. *Genome Biol.* **12,** R126 (2011).

155. Tan, G. C. *et al.* 5′ isomiR variation is of functional and evolutionary importance. *Nucleic Acids Res.* **42,** 9424-9435 (2014).

156. Liu, X.-D. *et al.* Overexpression of miR-34c inhibits high glucose-induced apoptosis in podocytes by targeting Notch signaling pathways. *Int. J. Clin. Exp. Pathol.* **8,** 4525–4534 (2015).

157. Bae, Y. *et al.* miRNA-34c regulates Notch signaling during bone development. *Hum. Mol. Genet.* **21,** 2991–3000 (2012).

158. Yu, F. *et al.* MicroRNA 34c gene down-regulation via DNA methylation promotes self-renewal and epithelial-mesenchymal transition in breast tumor-initiating cells. *J. Biol. Chem.* **287,** 465–473 (2012).

159. Yang, I. V. *et al.* Relationship of DNA methylation and gene expression in idiopathic pulmonary fibrosis. *Am. J. Respir. Crit. Care Med.* **190,** 1263–1272 (2014).

160. Cooper, G. *The Cell: A Molecular Approach*. (Sunderland (MA): Sinauer Associates, 2000).

161. Sanford, M. & Scott, L. J. Gefitinib: a review of its use in the treatment of locally advanced/metastatic non-small cell lung cancer. *Drugs* **69,** 2303–2328 (2009).

162. Ter Heine, R. *et al.* Fatal interstitial lung disease associated with high erlotinib and metabolite levels. A case report and a review of the literature. *Lung Cancer Amst. Neth.* **75,** 391–397 (2012).

163. Clark, J.-A. B. J., Tully, S. J. & Marshall, H. D. Sequence analysis of the Ras-MAPK pathway genes SOS1, EGFR & GRB2 in silver foxes (Vulpes vulpes): candidate genes for hereditary hyperplastic gingivitis. *Genetica* **142,** 517–523 (2014).

164. Yamazaki, T. *et al.* Role of Grb2 in EGF-stimulated EGFR internalization. *J. Cell Sci.* **115,** 1791–1802 (2002).

165. Rojas, M., Yao, S. & Lin, Y.-Z. Controlling Epidermal Growth Factor (EGF)-stimulated Ras Activation in Intact Cells by a Cell-permeable Peptide Mimicking Phosphorylated EGF Receptor. *J. Biol. Chem.* **271,** 27456–27461 (1996).

166. Tzouvelekis, A. *et al.* Increased expression of epidermal growth factor receptor (EGF-R) in patients with different forms of lung fibrosis. *BioMed Res. Int.* **2013,** 654354 (2013).

167. Han, F. *et al.* Emerging Roles of MicroRNAs in EGFR-Targeted Therapies for Lung Cancer, Emerging Roles of MicroRNAs in EGFR-Targeted Therapies for Lung Cancer. *BioMed Res. Int.* **2015,** e672759 (2015).

168. Yin, D. *et al.* miR-34a functions as a tumor suppressor modulating EGFR in glioblastoma multiforme. *Oncogene* **32,** 1155–1163 (2013).

CURRICULUM VITAE

[redacted]

[redacted]

[redacted]